SENSITIVITY ANALYSIS OF A COMBAT SIMULATION
USING
RESPONSE SURFACE METHODOLOGY

THESIS

William J. Berg, Major, USAF

AFIT/GOA/ENS/97M-01

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

19970429 229

SENSITIVITY ANALYSIS OF A COMBAT SIMULATION
USING
RESPONSE SURFACE METHODOLOGY

THESIS

William J. Berg, Major, USAF

AFIT/GOA/ENS/97M-01

THESIS APPROVAL
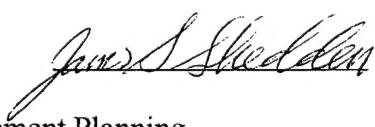
STUDENT: Major William J. Berg   CLASS: GOA-97M

THESIS TITLE: Sensitivity Analysis of a Combat Simulation Using Response Surface
Methodology

DEFENSE DATE: 13 February 1997

COMMITTEE:

NAME/DEPARTMENT   SIGNATURE

Advisor: Dr. Kenneth W. Bauer
          Professor of Operations Research
          Department of Operational Sciences, AFIT/ENS

Reader: Lt Col James S. Shedden
          Chief Analyst
          Aeronautical Systems Center, Director of Development Planning
          ASC/XRA

DISCLAIMER NOTICE
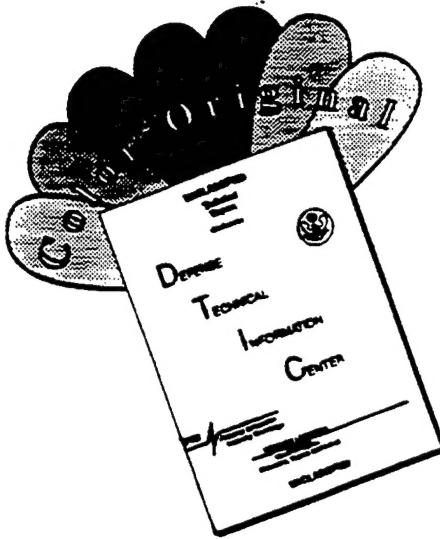
THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF COLOR PAGES WHICH DO NOT REPRODUCE LEGIBLY ON BLACK AND WHITE MICROFICHE.

AFIT/GOA/ENS/97M-01

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government

SENSITIVITY ANALYSIS OF A COMBAT SIMULATION
USING
RESPONSE SURFACE METHODOLOGY

THESIS

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Masters of Science in Operations Research

William J. Berg, B.S.

Major, USAF

MARCH, 1997

ACKNOWLEDGEMENTS

Table of Contents

List of Figures

List of Tables

AFIT/GOA/ENS/97M-01

Abstract

This thesis provides a methodology to analyze the sensitivity of a combat simulation relative to the capabilities of the modeled weapon systems. Response Surface Methodology (RSM) was the primary technique used for analysis. A ground battle scenario was developed with a commercial PC wargame *Empire II: The Art of War*™. In the computer verses computer mode, the wargame was transformed into a 'combat simulation'. A screening experiment was used to determine which of an original 23 weapon factors were actually significant to this scenario. After analyzing the remaining five factors, only three ended up as truly significant to the battle outcome. The three factors are used in a regression equation which predicts the outcome of the overall simulation scenario. The sensitivity of the simulation outcome is reflected by the coefficients of each factor. The regression model was then validated against independent simulation trials.

The method proved useful in analyzing the underlying behavior of a combat simulation. Combat analysts must isolate those weapon capabilities which have the greatest influence on the simulation model. Modeling funds can be concentrated on those factors. With accurate modeling and timely updates of those capabilities, the simulation model will earn more credibility as a decision aid.

SENSITIVITY ANALYSIS OF A COMBAT SIMULATION USING RESPONSE SURFACE METHODOLOGY

I. Introduction

Many combat simulations are used to compare various force structures for a given scenario. Typically, the number and type of combat units are manipulated to improve simulation outcomes. These outcomes are also influenced by hard-coded or "fixed" variables embedded into the simulation algorithm. Examples of these types of variables include the relative power of weapon systems, weather effects, and mobility factors. What happens to the simulation outcome when these variables are changed? Is the simulation sensitive or robust to these variables? Which are the most important? The answer to these questions can lead to a more credible simulation model. Significant variables and their characteristics should be isolated and periodically updated.

Objective

Apply Response Surface Methodology to analyze weapon system capabilities, and to identify those with the greatest influence on a combat simulation outcome.

Purpose

A simulation outcome, especially from an accredited model, has significant influence on force structure decisions. If an incorrect value of some significant factor can bias the

outcome, the decision maker will be working with incorrect data. Major decisions that involve lives and huge amounts of money need protection from this possibility. However, updating all the factors in a combat simulation can be a time consuming and expensive process. Such maintenance may have a tendency to be ignored. This technique will isolate the simulation's most significant factors. The simulation model may be updated and maintained in a much more economical manner.

Other benefits are also available from a response surface analysis of the simulation model. RSM will provide a metamodel of the combat simulation model itself. The simulation model can actually be verified and validated to some degree with the technique. For example, counter-intuitive behavior which may require further investigation could be revealed. Another major benefit of RSM is the quick-turn analysis opportunity of the resulting regression equation. The capability of instantly forecasting simulation results can be a valuable asset for any decision-maker.

Background

When a Commander-in-Chief (CINC) of a major Unified Command such as U.S. Central Command (USCENTCOM) is faced with a situation that may require military action in his Area of Responsibility (AOR), he is required to provide feasible Courses of Action (COA's) to the National Command Authority for consideration. Such was the case in 1990 when Iraq invaded Kuwait. General Norman Schwarzkopf (CINC-USCENTCOM) was tasked to provide a military response option. Some options were already in place but not for the exact scenario. Analysts were needed to compare options for possible conflict. Simulations ran against an expected enemy force resulted in

outputs for various allied force structures. These results were an aid in the decision of the number and types of units to be deployed to the Gulf region for Desert Shield.

The simulation used for this kind of support was TACWAR, a deterministic force on force tactical level simulation that USCENTCOM uses to support the Joint Operation Planning and Execution System (JOPES). This simulation allows looks at all kinds of force against force scenarios to find how various Operation Plans fare against an expected threat. TACWAR, like other simulations, has validated algorithms and is accredited for this kind of application.

However, these algorithms are a fixed value. They have fixed biases concerning various factors. No problem exists if the biases do not affect the simulation outcome, but over time they could. For instance, the offensive capability of a horse cavalry was relatively huge during the American Civil War. If computer war simulations existed at that time, an algorithm would have considered a unit with horses a high powered unit. In today's war, a unit with horses is a practically worthless unit. If the algorithm remained unchanged, this simulation would have become less credible over time.

To maintain a simulation model's credibility, an occasional review of the factors has to be considered. These factors need to be adjusted so to provide the most accurate result. Before these adjustments can be made, each factor must be measured for its effect in the field. This information can be gathered from testing at the proving ground, military exercises or expert opinion. With a limited budget, adequate information for all factors may be impossible to get. Using RSM techniques on the simulation model will isolate those most significant factors that need investigation. Specifically, an RSM

sensitivity analysis can highlight those factors that truly make a difference to the simulation outcome.

Altering the 'hard-coded' algorithms in TACWAR for a sensitivity analysis is infeasible at this thesis level. However, a similar type study can be made on a commercially available war game. *Empire II: The Art of War*™ is a computer war game that allows a player to change a number of the "fixed" variables such as strength values of weapon systems. Fortunately, this 'game' may be played in a computer verses computer mode, which essentially turns it into a combat 'simulation'. A successful analysis technique for this game most likely can be applied to an established simulation such as TACWAR.

Overview

The following report is split into five additional chapters. First, the Background chapter will cover the statistical foundation of RSM. The chapter will also cover the basic components of the *Empire II* wargame. A Literature Review will cover opinions and techniques offered by notable writers and previous theses authors. Chapter Four will cover the actual methodology used in the research. The rolling 'checklist' covers all the steps from start to finish. Results are developed and validated in Chapter Five. All data is presented in the appendices. Finally, the research and some of its benefits are reviewed with a couple of suggestions for follow-on research. The thesis hopefully will serve as a guide for those who have a similar objective, or as a foundation for those who are intent on expanding the technique.

II. Background

Introduction

This chapter explains the tools used in this research. First, an overview of Response Surface Methodology, otherwise known as RSM, will show how this regression process works. RSM will be used to find the factors most significant to a simulation outcome. The final RSM model will be subjected to several tests to determine if indeed the model is a reasonable representation of the simulation process. In this case, RSM will be used to analyze a combat simulation. The simulation is actually a commercially produced computer wargame that allows computer verses computer play. Wargames are generally considered combat simulations with human interaction, but in this application, only computers will be involved in the play of the game. An overview of the computer simulation is presented at the end of the chapter.

Response Surface Methodology -- RSM

Response Surface Methodology is collection of various statistical and mathematical techniques used to 'map' various processes by regression analysis (Myers and Montgomery, 1995: 1). RSM allows a process that allows numerous factors to be described with a single equation. Equation (1) describes a process with j factors.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_j x_j \qquad (1)$$

where
    x_j is the input factor's value.

    β_j is the coefficient or parameter that describes that factor's effect on the response.

    β_0 is the intercept value -- or the value of the response when all factors are zero.

The simplest form of (1) is the basic linear equation with a single factor:

$$y = mx + b \qquad\qquad (2)$$

where
  y is the response of the process.

  x is the input factor value.

  m is the coefficient that describe that factor's effect on the response (the slope).

  b is the intercept value -- or the value of the response when all factors are zero.

Since the response of a typical process generally follows a pattern based on input factors' values, some kind of 'mapping' of the process can be described with an equation. Mapping is a term to describe output values for various sets of inputs. For example, a surveyor can map the terrain of a plot of land by simply taking point elevation readings around the plot. By obtaining enough elevation readings around the plot, a picture, or a map, of the terrain begins to take form. Someone that has never seen the plot of land can get a good idea of what the land form is like. This is the basic premise of RSM.

Instead of taking elevation readings of every square inch of the plot, readings at each corner may supply enough information to map contours of the land. This is especially true if the land is known to be flat or gently slopes. But, if the relief of the land form is unknown or suspected of undulating, then many more readings throughout the plot are required to map the elevation contours accurately. By taking these readings in an organized fashion, say in a grid fashion, then a minimum number of readings are required. If this grid is laid out north-south and east-west, the resulting chart will have many elevation values laid out in an organized manner. So much so that one could pick a random location on the land plot, go to the chart and get a pretty good idea of what the elevation would be there.

A similarity can be drawn if the results of a process are equated with the elevation readings, while the factor values are equated with the coordinates, or location, of each reading. The amount of factor 1 is like the amount or distance traveled north from the southern edge of the land plot, while factor 2 is the distance east from the western edge. Go to the location on the map where the amount of factor 1 is so much (say 'A') and factor 2 is some amount (say 'B'), then a pretty good guess of the process outcome at that location can be made.

The beauty of RSM is that it allows many factors to be included when this map is built. Most industrial processes, from making steel to making pancake mixes, can be mapped out using this basic idea of RSM. Once the process is mapped out, if some desired output of the process is requested, go to the map and pick off the factor levels that generate the desired result. Industry will typically want the optimum result – no problem, go to the map and locate the peak!

If the process is a linear process, meaning the factors make the same amount of difference no matter where on the map one is looking, the optimum point will not only be on the edge of the map, but at one of the corners. Consider the land plot example. Where the plot is located on a side-slope of some ridge line, then the highest point will be along the highest edge. Unless the edge is parallel to the ridge line, the highest point on the plot will be on either end of the edge. The same reasoning goes for the lowest point. Therefore, the optimum point, whether at the highest or the lowest response, will be at a corner point (Figure 1).
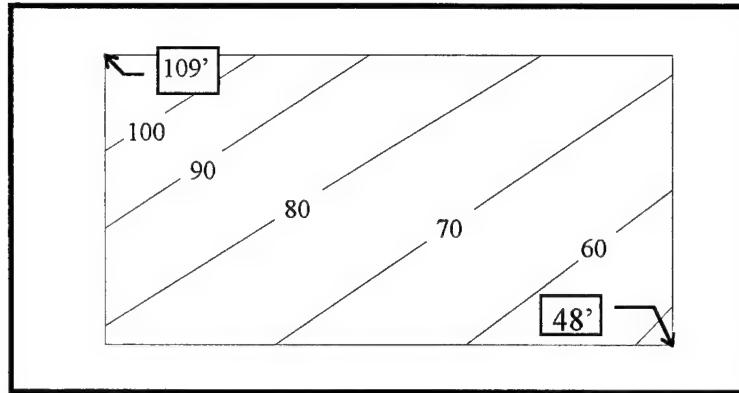
Fig. 1 -- Mapping of a Land Plot

However, most cases in industry involve non-linear processes. Consider the pancake

mix. If there is just a small amount of flour in the initial recipe, a unique result will

occur – lousy pancakes. Now, as the amount of flour is increased, the pancakes get

better, but only to a point. Once past that point, the pancakes worsen as the flour amount

increases. This is different from a linear process where theoretically pancakes would

only improve as flour is increased. Equation (1) above can only describe a linear

process.

Another related but slightly more complex equation can handle a non-linear process:

$$y \ = \ \beta_0 \ + \ \beta_1 x_1 \ + \ \beta_2 x_2 \ + \ \beta_{12} x_1 x_2 \tag{3}$$

Like the original equation, equation (3) is considered a 'first order' equation because

there are no quadratic , cubic, etc. terms. But to describe the flour and the pancake mix

situation, squaring the flour factor value will describe the effect of getting better then

worse pancake results. An equation that can handle two factors and their square terms is written:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 \qquad (4)$$

Equation (4) is known as a second order equation because of its quadratic terms. Second order equations can get quite complex, but they do a good job mapping the undulating surface of responses. A linear equation can only describe a flat surface, which is inadequate for the more typical process whose response behaves in such a way that there are 'peaks' and 'troughs' based on various factor inputs. While second order equations can become very hard to work with, a calculator or computer can rapidly produce a result!

When optimizing the result for some given factors in a non-linear case, calculus offers the derivative for finding the maximum of minimum. Looking at equation (4) and taking the partial derivatives against x_1 and x_2 results in:

$$\frac{\partial y}{\partial x_1} = \beta_1 + 2\beta_{11} x_1 + \beta_{12} x_2 \qquad (5)$$

and

$$\frac{\partial y}{\partial x_2} = \beta_2 + 2\beta_{22} x_2 + \beta_{12} x_1 \qquad (6)$$

By setting both equations to equal zero, then a unique solution for x_1 and x_2 will result. This solution are the factor values for the optimum response y. Larger models with k factors will require k equations to solve. These can be arranged using matrices, and solved by taking the Jacobean or derivative of the matrix and setting them equal to zero. Computer programs can make this process relatively painless.

Rapidly calculating a result is one of the great benefits of RSM. RSM is a tool that mathematically models or 'maps' a process. With this map, a credible estimate of an output can be generated from a given set of inputs. The output can be optimized by using calculus Max / Min procedures on the model equation. The next two sections will show how the equation is built.

Where Does the Regression Equation Come From?

Assuming that a process will have exactly one possible result for a given set of inputs, and responses have been collected for a number of input combinations, then a RSM model may be constructed. In a case where there is one factor, two settings that result in two outputs could simply be plotted as two points on graph paper. An RSM model of the process is produced by connecting the two points with a line. At any particular factor level, the corresponding point on the line is a prediction of the response. This is the best model that can be constructed from the given information. But, it may be risky to claim that a model based on two observations is an adequate model.

Say by taking 10 observations, each at a unique input value, results in a series of plots that form a trend. Taking a straight edge and pencil, one can draw a line that pretty well fits the trend. This is a form of a regression line. Compared to a two-point model, a regression line is a more dependable at describing the process than the two point model. A single factor is the limit of this method since the one factor and its response requires a two dimensional surface to model. Two factors and their response would need a three dimensional plot, while k factors will need $k + 1$ dimensions to model.

A mathematical approximation method process called the least-squares fit, puts a line through the plot of points such that the cumulative distance between the line and the actual plots is minimized. For mathematicians, that is not quite the case. Since some points are above the line, they would have a positive distance from the line, while the ones below would have a negative distance. By squaring each distance value, all resulting values (distance squared) will be positive. Adjusting the line to minimize this cumulative value will result in the 'best' fit line. Therefore, regression equation (6) used in RSM will have some error although the error is minimized:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_j x_{ij} + \varepsilon_i \qquad (6)$$

where
$i = 1, 2, 3, \ldots, n$ (n plotted points)
$j = 1, 2, 3, \ldots, k$ (k factors)

ε_i is the individual error between plotted point i and the regression equation.

Squaring each ε_i and summing gives the total sum of squares (L), which is to be minimized.

$$L = \sum_{i=1}^{n} \varepsilon_i^2 \qquad (7)$$

With multiple factors (x_{ij}) and the various responses (y_i), the equation can be simplified into vector and matrix form.

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad (8)$$

11

where

$\quad$ **y** $\quad$ is the response vector.

$\quad$ **X** $\quad$ is the factor matrix.

$\quad$ β $\quad$ is the coefficient vector.

$\quad$ ε $\quad$ is the error vector.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots x_{1k} \\ 1 & x_{21} & x_{22} & \cdots x_{2k} \\ 1 & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots x_{nk} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The values for **y** and **X** are known. The best fitting equation will have the coefficient vector β. To find an estimate for vector β, an estimator called vector **b** is used. Vector **b** is the defining vector for the RSM regression model. Solving for **b** is the basic problem in RSM. This can be done by minimizing L where:

$$L = \sum_{i=1}^{n} \varepsilon_i^2 \quad \text{or} \quad L = \varepsilon'\varepsilon \tag{9}$$

Since

$$\varepsilon = (\mathbf{y} - \mathbf{X}\beta) \tag{10}$$

then

$$L = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \tag{11}$$

Multiplying out

$$L = \mathbf{y}'\mathbf{y} - \beta'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \tag{12}$$

$\beta' \mathbf{X'} \mathbf{y}$ is a scalar value whose transpose has the same scalar value, therefore

$$L = \mathbf{y'} \mathbf{y} - 2\beta' \mathbf{X'} \mathbf{y} + \beta' \mathbf{X'} \mathbf{X} \beta \qquad (13)$$

Let vector **b** be the estimator for β. The least square estimator must satisfy:

$$\frac{\partial L}{\partial \beta}\bigg|_{\mathbf{b}} = -2 \mathbf{X'} \mathbf{y} + 2 \mathbf{X'} \mathbf{X} \mathbf{b} = \mathbf{0} \qquad (14)$$

Simplify

$$2\mathbf{X'}\mathbf{X}\mathbf{b} = 2\mathbf{X'}\mathbf{y} \qquad (15)$$

then

$$\mathbf{X'}\mathbf{X}\mathbf{b} = \mathbf{X'}\mathbf{y} \qquad (16)$$

Thus **b**, the least squares estimator of β, reduces to:

$$\mathbf{b} = (\mathbf{X'}\mathbf{X})^{-1} \mathbf{X'} \mathbf{y} \qquad (17)$$

So, by supplying the **X** matrix and the **y** vector, the best fit least-square regression equation can be found:

$$\hat{\mathbf{y}} = \mathbf{X} \mathbf{b} \qquad (18)$$

where $\hat{\mathbf{y}}$ is the estimate for the response (or the model's response) for any combination of input factors (**X**). (Myers and Montgomery, 1995: 19-21)

Unfortunately, supplying the **X** matrix and **y** vector is a time consuming process. This equates to the land surveyor going out and getting all the point elevations on the land plot. An organized process of gathering data for the equation is required. Design of Experiments offers some very efficient methods for doing this.

Design of Experiments

The goal of an experiment is to provide adequate data about a process for use in the construction of credible models. These models can then provide insights of the process. An efficient experiment accumulates this information with minimal observations. Each experimental trial increases data. However, there is a point where the additional information does not justify the cost of an additional trial. Thus, a good design of experiment will get adequate data at a reasonable cost. This section will develop the reasoning behind the experiment selected for this research.

A thorough experiment will consider each and every possible combination of all possible factors in a given scenario. For instance, if six factors influence a certain response, each individual factor would be metered at certain levels and then examined at each level. If each factor is separated into say five levels, then a thorough experiment will have 5^6 or over 15,000 trials! If only two levels are considered (hence a Two-Level Factorial experiment) there would only be 2^6 or 64 trials. (Myers and Montgomery, 1995: 79)

A Two-Level Factorial experiment uses only a high and low value for each factor. For simplicity, each high and low value is 'coded' as 1 and -1. Median values, if needed, are coded as zero. The experimenter selects the high and low value of each factor to code. For instance, if one of the factors under consideration is temperature, the experimenter could have a region of interest from $32°$ to $58°$. This can be translated as $45° \pm 13°$.

14

A general formula for coding:

$$x_i = \frac{\xi_i - \xi_{i0}}{S_i} \qquad (19)$$

where

x_i is the coded value (-1, 0, 1)

ξ_i is the uncoded low, median, or high value

ξ_{i0} is the uncoded average value

S_i is the difference between the uncoded median and high (or low) values.

(Box and Draper, 1987 :107)

In this example the high value would translate to '1' or $\quad 1 = \dfrac{58-45}{13}$

Some experiments will call for some coded factors values in excess of '1'. For instance, if this experiment called for a value of '2', the actual value called for can be calculated:

$$2 = \frac{\xi_i - 45}{13} \quad \text{or} \quad \xi_i = 71° \qquad (20)$$

This coding scheme allows factor values to fit into even more advanced Two-Level experimental designs. These Two-Level designs provide a specific **X** matrix which radically reduces the number of experimental trials required for credible results, although in the end the conclusions must be decoded again!

Consider the six factor Two-Level experiment. Compared to the 5^6 experiment, some information will be lost, but not enough to justify 15,000 more trials! Even 64 (2^6) trials may be too many. If trials are expensive, any way to cut the number of trials should be

15

considered. Enter the <u>Two Level **Fractional** Factorial</u> experiment design. Of the 64

trials, many measure interactions of three or more of the six factors. Rare are the cases

where a certain combination of three or more factors make a significant difference on the

outcome. According to Dr. Genichi Taguchi, an expert in Design of Experiments, even

two-factor interactions tend to be insignificant (Taguchi, 1987: 3). The fractional

factorial design essentially eliminates those trials that measure multiple interactions. For

instance, if the experimenter can only afford 20 or so trials, then a 16 trial design can be

used. Granted, many of the high level interactions will be lost, but all the main effects

will be preserved. That should be the main goal of the experimenter.

A Resolution III experiment design is a design that has been reduced in such a way

that in some cases it is unclear which had an impact on the response – a main factor or a

two-factor interaction. This is called an 'alias' relationship. Aliases are the cost of

reducing the number of experiments. However, the combinations of factors in each trial

can be managed such that the aliases are minimized if not eliminated. If all aliases

between main factors and two-factor interactions are eliminated, then the design is

referred to as a Resolution IV experiment. A Resolution IV experiment can be attained

in twice the trials of any Resolution III experiment by simply augmenting with an

identical second experiment, except for reversing the values of each factor. This is called

a fold-over design. The fold-over is an option if it is unclear if either a main factor or

one of the interactions is causing the response. If the offending interaction is known,

then it is possible to manipulate some Resolution III experiments to eliminate that

particular alias. This alternate type of experiment can get the information of a

Resolution IV experiment form the same number of trials of a Resolution III experiment! Experimental design documents will have tables to aid those designs. (Wu and Chen, 1992: 162-175)

There will be at least one pair of two-factor interactions aliased with each other in a Resolution IV experiment. If all two-factor interactions are of interest, then even a Resolution IV experiment will be insufficient. The outcome of the experiment may be obviously dependent on an interaction of two factors, but when two of these interactions are aliased with each other, we can't really conclude which interaction caused the response. A Resolution V experiment is required. A Resolution V experiment deconflicts all two-factor interactions with one another. The Resolution V design is generally much more expensive because it requires at least double the trials of the Resolution IV experiment. Nevertheless, if two-factor interactions are important, the Resolution V design may be worth the extra cost.

Whatever design is selected, it defines the **X** matrix. The responses for each of the trials in the experiment defines the **y** vector. Refer to the six factor experiment mentioned. Figure 2 shows an eight trial Resolution III experiment.

	Factor x1	Factor x2	Factor x3	Factor x4	Factor x5	Factor x6	Response
Trial 1	-1	-1	-1	1	1	1	y_1
Trial 2	1	-1	-1	-1	-1	1	y_2
Trial 3	-1	1	-1	-1	1	-1	y_3
Trial 4	1	1	-1	1	-1	-1	y_4
Trial 5	-1	-1	1	1	-1	-1	y_5
Trial 6	1	-1	1	-1	1	-1	y_6
Trial 7	-1	1	1	-1	-1	1	y_7
Trial 8	1	1	1	1	1	1	y_8

Fig. 2 – Six Factor Resolution III Design

Figure 3 reflects the corresponding to **X** matrix and **y** vector.

$$
\mathbf{X} = \begin{bmatrix}
1 & -1 & -1 & -1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & 1 & 1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\qquad
\mathbf{y} = \begin{bmatrix}
y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8
\end{bmatrix}
$$

Fig. 3 – Resulting **X** matrix and **y** vector

The first column of the **X** matrix is all ones and is referred to as the x_0 column. This column corresponds with b_0 of the **b** vector while b_{1-6} corresponds with each of the six factors (x_{1-6}).

One of the advantages of this particular design is that it is orthogonal. Multiply each element of one column against the corresponding element in any other column. The sum of the resulting products will always be zero. The advantage of this is that $\mathbf{X'X}$ is a diagonal matrix which is much easier to handle if doing this calculation manually. Another advantage of an orthogonal design that it provides the experimenter the option to isolate a factor. A change in a single factor unilaterally affects any corresponding difference in the response.

Resolution IV experiment form the same number of trials of a Resolution III experiment! Experimental design documents will have tables to aid those designs. (Wu and Chen, 1992: 162-175)

There will be at least one pair of two-factor interactions aliased with each other in a Resolution IV experiment. If all two-factor interactions are of interest, then even a Resolution IV experiment will be insufficient. The outcome of the experiment may be obviously dependent on an interaction of two factors, but when two of these interactions are aliased with each other, we can't really conclude which interaction caused the response. A Resolution V experiment is required. A Resolution V experiment deconflicts all two-factor interactions with one another. The Resolution V design is generally much more expensive because it requires at least double the trials of the Resolution IV experiment. Nevertheless, if two-factor interactions are important, the Resolution V design may be worth the extra cost.

Whatever design is selected, it defines the **X** matrix. The responses for each of the trials in the experiment defines the **y** vector. Refer to the six factor experiment mentioned. Figure 2 shows an eight trial Resolution III experiment.

	Factor x1	Factor x2	Factor x3	Factor x4	Factor x5	Factor x6	Response
Trial 1	-1	-1	-1	1	1	1	y_1
Trial 2	1	-1	-1	-1	-1	1	y_2
Trial 3	-1	1	-1	-1	1	-1	y_3
Trial 4	1	1	-1	1	-1	-1	y_4
Trial 5	-1	-1	1	1	-1	-1	y_5
Trial 6	1	-1	1	-1	1	-1	y_6
Trial 7	-1	1	1	-1	-1	1	y_7
Trial 8	1	1	1	1	1	1	y_8

Fig. 2 – Six Factor Resolution III Design

The design has given the experimenter an efficient way to come up with both an **X** matrix and **y** vector – the ingredients required to make a regression model. But, with the efficiency offered by the fractional experiment design, some accuracy is lost. To get an idea of how effective the model is, some tests have to be made. If these tests show that the resulting regression model does not fit the data well, then a supplemental experiment may be required to increase data points. This additional information will allow the construction of a more complex but better fitting regression model. Some of these tests are discussed in the next section.

Measure of Goodness of the Regression Equation

After running the experiment, the **X** matrix and **y** vector are known values. The **b** vector can now be solved for using equation (17). How well does the **b** vector estimate β ? Five tests are used in this research to determine if the fit is good or not. These include the Coefficient of Multiple Determination (R^2) analysis, residual analysis plots, the 't' ratio test, 'F' test for Regression Relation, and if needed, the Lack of Fit test (Neter, 1996: 229). The final determination of model fit tends to be a subjective one, but these test results are used to defend the judgment call.

First, look at the elements that make up these tests. A simple one factor (x) model regressed from four data points (Figure 4). (Neter, 1996: 70)
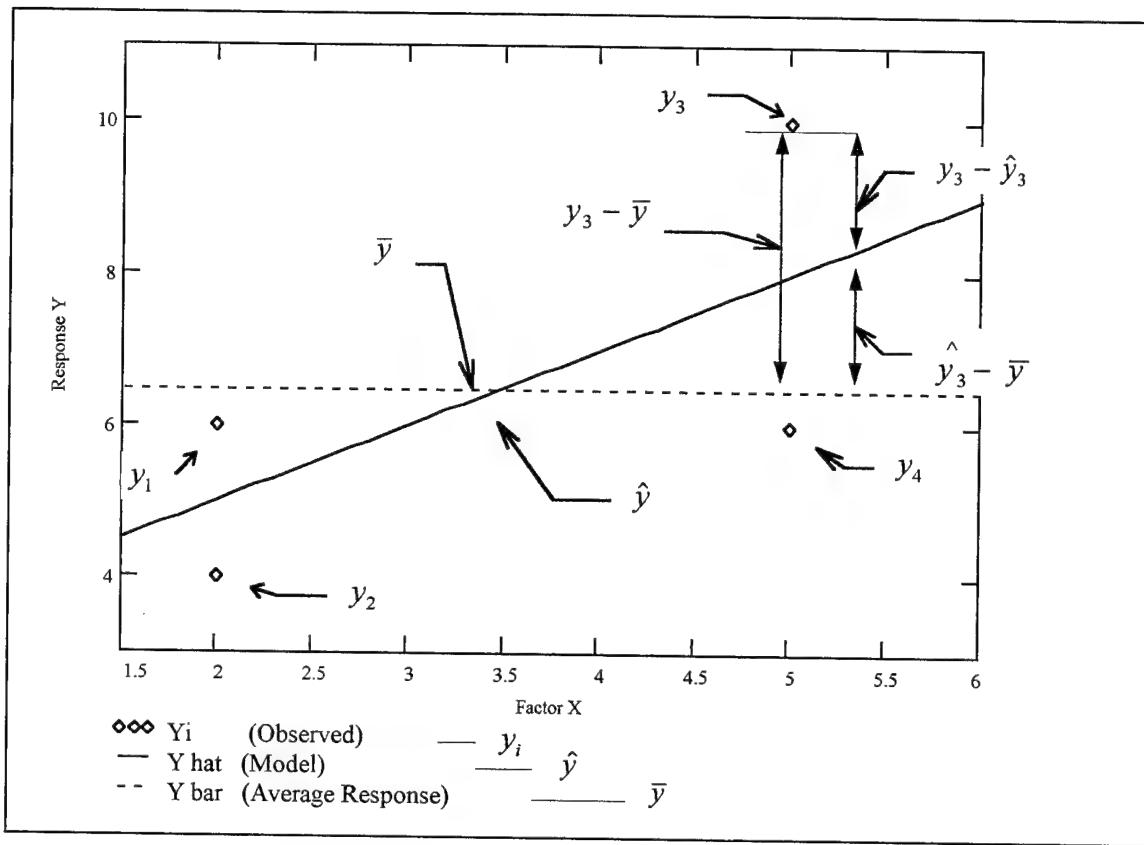
Fig. 4 – Terms of a Regression Plot Defined

Some definitions:

Regression Sum of Squares -- $SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2$ (21)

Error Sum of Squares -- $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ (22)

Total Sum of Squares -- $SSTO = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$ (23)

These definitions have the relationship:

$$SSR + SSE = SSTO$$ (24)

This can be interpreted using Pythagorean Theorem -- $a^2 + b^2 = c^2$ (Figure 5)
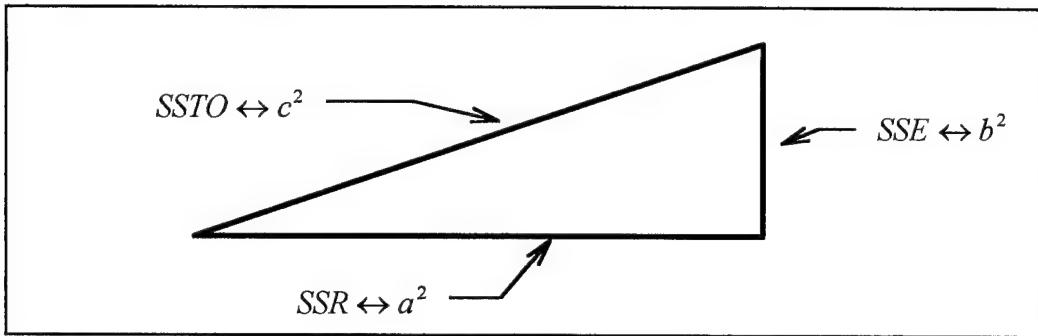
Fig. 5 – Relationship Between Error and Pythagorean Theorem (RSM Class Notes)

<u>Coefficient of Multiple Determination</u> (R^2). The diagram in Figure 5 is a good picture of what the Coefficient of Multiple Determination, better known as R^2, is measuring. R^2 is the amount of variance that the model explains. This is the proportion of SSR to $SSTO$. (Neter, 1996: 230)

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \tag{25}$$

Generally, the closer R^2 is to 1.0, the better. A model that explains a response to a single factor with only two data points can be interpreted as a great model. In this case R^2 will have a perfect value of 1.0. However, since only two data points exist, a regression model would simply connect them. The model would have zero error since both observed data points fall on the model's plotted line. All variance would then be explained and R^2 would be a perfect 1.0. This model would be questionable at best, since so few data points are used to define it. In higher dimensions, the same problem exists.

In Plackett-Burman designs, just enough trials are used to build a regression model. For instance, a 24 trial experiment can model a response dependent on 23 factors. The same situation occurs as in the two data point model, but in this case all 24 data points easily connect in 23 dimensions, and R^2 is 1.0! Although this kind of design can be used to get an idea of which factors are important and which are not, this model is not well suited to describe any response accurately. Thus a Plackett-Burman design is used just for screening out insignificant factors. (Kleijnen, 1975: 331)

If the R^2 value is relatively low, say 0.6 to 0.9, the model is considered only a fair fit. One of three reasons may explain for this kind of reading. First, the model simply does not describe the data. If this is really the case, the R^2 value will generally be lower than 0.6. But, a low R^2 could also be because of a second reason. If trials are repeated at each design point, and a wide variance of responses are recorded, a regression model can only place a single line through the band of data points. The distance between each data point and the regression line is squared and summed to give a relatively large SSE and in turn a smaller SSR. Therefore, R^2 is driven downward. The model is a good fit, but the wide variance puts the model in question. This is a common problem in modeling any process with large variability. Modeling processes such as economic behavior, business simulations, and combat simulations are but a few examples.

The Residual Plot. A third reason that R^2 may have a low value is because the data may have a non-linear pattern. A linear regression model will not adequately describe the data. One of the most important tests for detecting this is quite a subjective one. The residual analysis graph is simply a plot of the errors from the regression model. The

regression model is a best guess of the response. Each error is simply $y_i - \hat{y}_i$. Since the regression model plots through the set of data points, nearly the same number of positive and negative errors should occur. A good residual plot will have no discernible pattern. That is the subjective nature of the test. Figure 6 shows a lack of pattern while Figure 7 shows a parabolic pattern indicating that some kind of quadratic term is needed in the regression model. (Myers and Montgomery, 1995: 42)
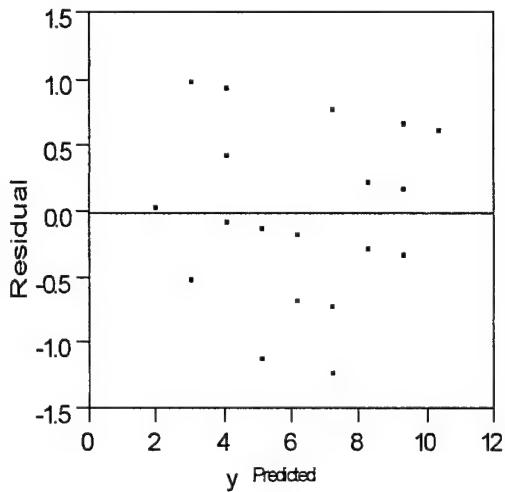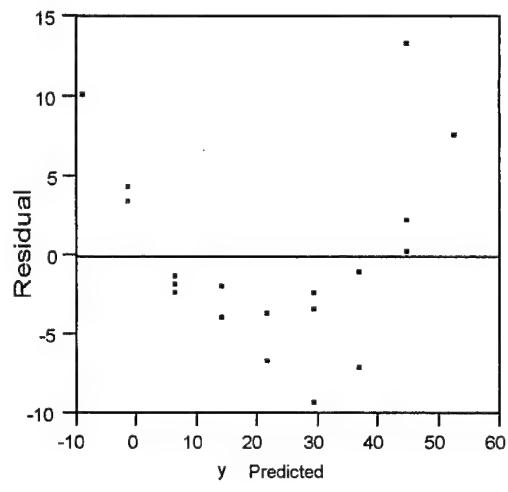
Fig. 6   Fig. 7

Residual Analysis Plots

The residual plot offers a simple visual check of the regression model. However, the plot is one of the most important tests to check the adequacy of the regression fit. The residual plot should always be considered regardless how good R^2 is.

The other three tests are more statistical in nature. The 't' ratio test is designed to test the significance of each coefficient b_i of the regression model. The 'F' test will show the overall significance of the entire model as a whole. Finally, the Lack of Fit test statistically does what the residual plot did visually. This test will determine if the

regression model is a good enough fit despite any non-linear effect. All these tests depend on the normality and the independence of residual error.

The 't' Ratio Test. Before any factor of a model can be declared significant, each must pass the '*Student t*' or 't' ratio test. The 't' test is a statistical tool that determines if a factor is indeed significant and not just noise. First, assume a factor has no effect on the response, or in other words, a regression model should assign it a coefficient of zero. After running an experiment, the data may seem to indicate that the outcome is indeed influenced by a particular factor. The simulation outcome may show some 'tendency' of dependence on a certain factor. However, if the tendency appears to be minor, the regression model really cannot claim the factor has a significant influence on the outcome. The factor is significant only if the tendency of the outcome is more pronounced than any randomness of the process. Standard error is a measure of randomness. The 't' test is a look at the ratio of the change of a process tendency to the standard error. (Myers and Montgomery, 1995: 31-32)

To illustrate the 't' test, assume some process has a tendency to output a value of y_0 with standard error $s\{y\}$. After changing some factor x_1 the process has some new tendency to output a value of y_1. But, before a model can 'claim' that x_1 really made the difference, a 't' test must be performed. First, the 't' ratio is calculated:

$$t = \frac{y_1 - y_0}{s\{y\}} \qquad (26)$$

where standard error $s\{y\} = \sqrt{\hat{\sigma}^2 C_{ii}}$.

C_{ii} is the diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ corresponding with the coefficient b_i of factor i, or in this case, C_{11}. A value of 't' is generated. Before any conclusions can be made, 't' is compared with a reference value $t*$. The value of $t*$ is based on:

1. The confidence required for the claim (termed as 'x percent confident').
2. The amount of data to support the claim (known as 'degrees of freedom').

If $t > t*$, then the claim is accepted. This is known as the Test of Hypothesis. The null hypothesis H_0 corresponds to the statement "The change from x_0 to x_1 makes no difference", while the counter hypothesis H_a says the change did. The calculated value 't' is the 'test statistic' which is compared against $t*$ resulting in one hypothesis favored over the other. Typically, a 95% confidence level is required before a null hypothesis is 'overturned'. The higher the confidence level, the higher $t*$, and the tougher it is to 'overturn' H_0.

The amount of data to support a claim is also important to the t^* value. One data point cannot show a tendency, but two can. Three points can show a better tendency, and 20 points even better. The term 'degrees of freedom' describes this fact. A claim can be made based on one data point, but the other data points support the claim. If 20 data points are available, 19 can be used to support the claim. Some hypotheses make several claims as will be shown in the explanation the 'F' test. In a case where 7 claims are made, then only 13 of the 20 data points remain for support. Degrees of freedom is simply the number of data points minus the number of claims. With fewer degrees of freedom, the higher $t*$ is.

With a known confidence level, and a known number for the degrees of freedom (df), $t*$ can be found in statistical tables. As the df exceeds 30, then the Central Limit Theorem states that df becomes an insignificant input. In the case where df exceeds 30 and 95% confidence is required for the claim, $t*$ will have a value of 1.96. This translates into saying that the difference between y_0 and y_1 must exceed 1.96 times $s\{y\}$ before experimenter can claim a difference. To illustrate, refer to Figure 8.

Plots Reflecting the Tendancy of Two Responses

y_1 central tendency is 2.5

The central tendency of y_1 exceeds 1.96 standard deviations of the y_0 tendency, therefore can be 'claimed' as distinct with 95% confidence

± 1.96

— Prob y_0 – – Prob y_1
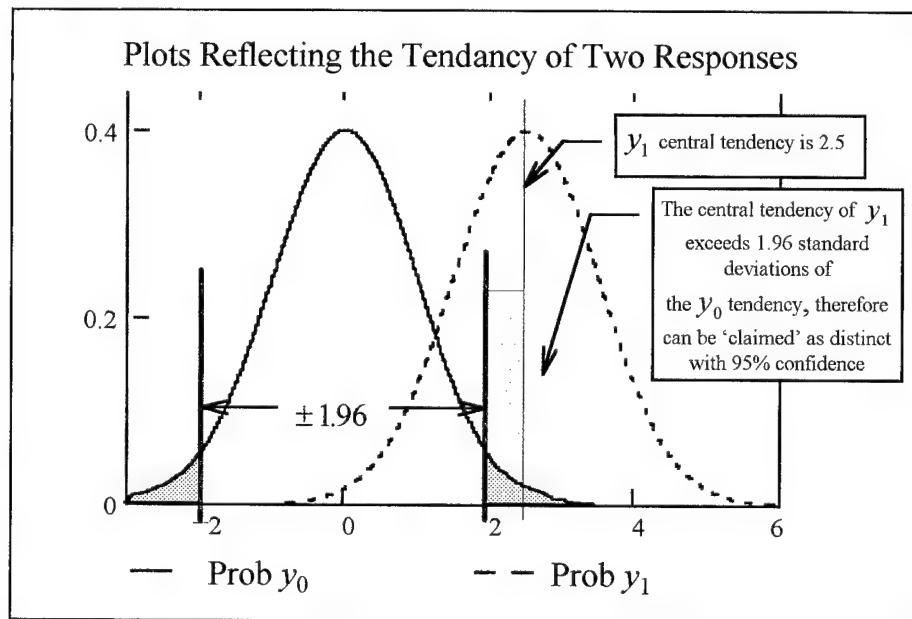
Fig. 8 – The Difference Between Factor x_0 and x_1

Overall, the 't' ratio test shows _if_ there really is a statistical difference between two response values. If so, the factor that influenced the difference can be declared significant at the confidence level defined. Three items influence the behavior of the 't' ratio test. It is harder to show a difference between two responses if:

1. The variance (standard error) is high.
2. There is a shortage of data to support the claim.
3. A high confidence level is required for the claim.

Commonly, the 't' ratio test will be put in terms of confidence level. For whatever 't' calculated, an equal t^* is determined from the existing df and then varying the confidence level. Generally, the confidence level is in terms of $(1 - \alpha)$, where α is the risk of a wrong claim. A wrong claim is referred to in hypotheses testing as a 'Type I' error. In making any claim, the researcher specifies an acceptable α, or probability of a Type I error. In the example above, α was specified at 5%, and the associated t^* was 1.96. If a number of factors are being considered at one time, many computer products will provide a 't' ratio for each of them. Significant factors are recognized by inspection when their associated 't' values exceed 1.96. An alternate method displays a probability value that could be compared against the specified α. These values are found in the Prob > |t| column. A value smaller than the specified α will reject the null hypothesis, and the factor is significant. On the other hand, if the value is higher than α, then the null hypothesis stands, and the factor is rejected as insignificant. This allows the researcher to recognize and disregard any non significant factors by simple inspection. Refer to Figure 9. Assume a 95% confidence level is required before claiming a factor's significance, or that the specified α is 5%. Notice that X1 and X4 have probabilities that exceed α, therefore both X1 and X4 are excluded from further consideration.

```
┌─────────────────────────────────────────────────────────────────┐
│   ╭──────────────────────╮                                        │
│   │  Parameter Estimates  │                                       │
│  ╭─┴──────────────────────┴──────────────────────────────╮       │
│  │ Term          Estimate    Std Error    t Ratio   Prob>|t|      │
│  │ Intercept     157.23333    3.803367     41.34    <.0001        │
│  │   X1          -1.708333    4.252293     -0.40    0.6913 ◄──────│
│  │   X2          14.958333    4.252293      3.52    0.0017        │
│  │   X3          42.291667    4.252293      9.95    <.0001        │
│  │   X4          6.9583333    4.252293      1.64    0.1143 ◄──────│
│  ╰───────────────────────────────────────────────────────╯       │
└─────────────────────────────────────────────────────────────────┘
```

Fig. 9 – Factor Coefficients and Their Respective '*t*' Test Result

The Significance of Regression Test.   A test for significance of regression is a test to determine if any linear relationship exists between a group of factors $(x_1, x_2, x_3, \ldots, x_i)$ and a response y. (Myers and Montgomery, 1995: 34)

The two hypotheses are:

$$H_0: \quad \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_i = 0 \tag{27}$$

$$H_a: \quad \text{at least one } \beta_i \neq 0 \tag{28}$$

The 'F' test, similar to the '*t*' ratio test, is used to determine if a model is really capable of claiming <u>any</u> factor significance.   The test statistic 'F' is a ratio that uses SSR, SSE, and their respective degrees of freedom. An Analysis of Variance (ANOVA) table describes the calculation of 'F' (Figure 10).

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F
Regression	SSR	k	$MSR = \dfrac{SSR}{k}$	$\dfrac{MSR}{MSE}$
Residual	SSE	n-k	$MSE = \dfrac{SSE}{n-k}$	
Total	SSTO	n		

Where: n is the total available degrees of freedom (Total data points - 1)

k is the number of factors

Fig. 10 – An ANOVA Table

Like the 't' ratio test, the test statistic 'F' is compared against an $F*$. $F*$ is a value determined by the degrees of freedom of both MSR and MSE, and the desired confidence level. If 'F' exceeds $F*$, then the researcher can claim some factor x_i is significant. Figure 11 reflects an ANOVA table with the addition of a probability value to compare against a specified Type I error probability (α). In this case, at least one x_i is significant.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	49528.167	12382.0	28.5322
Error	25	10849.200	434.0	Prob>F
C Total	29	60377.367		<.0001

Fig. 11 – 'F' Test Result

The '*t*' test and the 'F' test are very similar in nature. A claim about a model's behavior can be supported by using these tests. Typically, the 'F' test is first used to see if any factor is significant. If so, then each factor is examined by the '*t*' test. These are the main statistical tools to determine which factors are important in a process. In this research, these are also the main tools for determining the sensitivity of a battle simulation's outcome to various input factors.

The Lack of Fit Test. The last statistical test before a model is considered usable is the Lack of Fit (LOF) test. Like the residual plot, the LOF test is used to determine if the regression fit indeed models the process adequately. The statistical LOF test is much more sensitive to a bad model than the subjective residual plot. If multiple trials are made at the factors' median values (coded as 0 in the experiment), then a LOF test can be performed. In this case the null hypothesis says the model fits the process. An 'F' test statistic is also used in the LOF test. With repeat observations of some trials, SSE can be separated into two components, SS Pure Error (SSPE) and SS Lack of Fit (SSLOF).

The degrees of freedom for SSPE is the total number of replicates in the experiment. The ANOVA table in Figure 12 explains how the LOF 'F' statistic is calculated. (Myers and Montgomery, 1995: 50)

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F
Regression	*SSR*	*k*		
Residual	***SSE***	***n-k***		
Lack of Fit	***SSLOF***	***c-k***	$MSLOF = \dfrac{SSLOF}{c-k}$	$\dfrac{MSLOF}{MSPE}$
Pure Error	***SSPE***	***n-c***	$MSPE = \dfrac{SSPE}{n-c}$	
Total	*SSTO*	*n*		

Where: c is number of trial points, and k is the number of factors

Fig. 12 – Lack of Fit ANOVA Table

With a good fitting model, the null hypothesis will hold. The calculated 'F' statistic should be less than $F*$. The probability value should exceed the specified Type I error (α). This is exactly the opposite of the previous 'F' test! Figure 13 indicates the regression model fits the process adequately.

Lack of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack of Fit	4	2322.533	580.633	1.4300
Pure Error	21	8526.667	406.032	
Total Error	25	10849.200		

Prob>F
0.2589

Fig 13 – Lack of Fit Result

A Final Quality Test. A final test, before making any declarations, is the sanity check. A regression model may claim a factor behaves in a counter-intuitive manner. Although this is another subjective test, the sanity check may do the most for the model's credibility. For instance, a regression model indicates that as a factor's value is increased, the response significantly decreases. However, common knowledge is that as the factor value increases, the response increases. Another possibility is that an unexpected factor seems to over-influence a response. As the factor value changes, the response reacts in a way that is too good to be true. Either way, an investigation is needed.

The least expensive investigation is simply retrace the experiment to see if data was input correctly. If other factors similar to the one in question are available, do they behave in a similar manner? If these bring no conclusions, then a confirmation experiment is in order. Isolating the factor in question by varying it while holding all others constant, may confirm or reject the sanity check. A possibility exists that the model may have discovered something unexpected! Either way, an unexpected model behavior has to be acknowledged and investigated. The credibility of the experimenter will be maintained if not enhanced!

Empire II: The Art of War™ Description

Empire II is a computer wargame designed for the entertainment of wargame enthusiasts. It was designed to allow the wargamer a opportunity to custom build scenarios to either play against other gamers or a computer. An option for computer against computer is also available. In this thesis, the computer verses computer option is used to avoid the long drawn out trials of human play, not to mention the complex experimental blocking designs due to several players. By taking the human player out of the process, *Empire II* is really no longer a wargame, but a combat simulation. Ironically, *Empire II* is not unlike some of the older combat simulations used in the military.

Movement. Like a typical manual wargame, *Empire II* is played on a regular 'game board' that allows two sides to play out a tactical level battle scenario. Unlike most manual wargames, *Empire II* offers many different 'game boards', simply by allowing the player to custom build them. (New World Computing--User Manual, 1995: 78-116) All *Empire II* game boards consist of a regular network spatial structure of square nodes (Figure 14). Each of these squares can be assigned one of 12 terrain types. Like any regular spatial network, all of the nodes (squares) are of one size. Entity movement from node center to node center (arcing) is allowed in any of eight directions throughout the network. Each node is assigned a terrain type that influences the movement of entities. A node with a terrain type 'charges' an entity a certain amount of 'offensive potential' to transit. The modeler has the freedom of assigning just how restrictive each

33

terrain type is for the movement of each entity type. A node may be assigned a terrain

type (such as swampland) that restricts movement. Arcing through these nodes will have

a high cost. Other nodes (such as clear terrain) will allow relative freedom of movement.
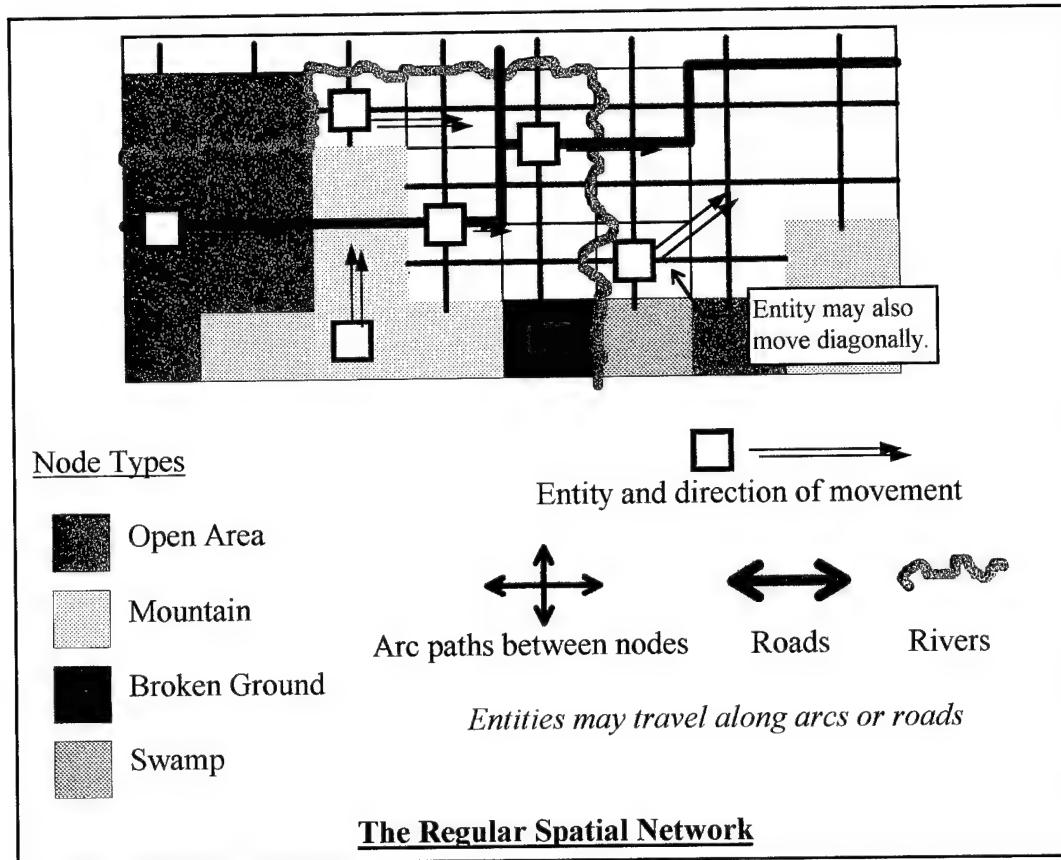
Fig. 14 – Entity Movement in *Empire II: The Art of War*™

There is another option that will allow for entity travel. Lines of communication such

as roads may be constructed along node edges. There is still a cost per unit length for

entity movement along these lines, but the higher expense of arcing from node center to

node center can be avoided. On the other hand, these edges between nodes can also

serve as a barrier between nodes (like rivers). In this case, as an entity arcing from node

to node crosses one of these, an additional cost is assessed. This has the effect of slowing down the movement of entities when obstructions are encountered.

Entities are allotted a fixed value of offensive potential for each successive turn. The game refers to this value as *movement* points. This potential may be used for either movement or for firing weapons (which are discussed later). The entity may be commanded to move towards a destination or stand-fast. The commander determines the best path through the node network. By arcing through low cost nodes or along lines of communication, the entity can reach its destination sooner with the least loss of offensive potential. Any remaining movement points may be used for offensive weapons.

The entities themselves represent the battle units. Each entity type is generic until defined by the modeler. Each entity type is assigned an amount of offensive potential in accordance to the type of weapon system it represents. Since offensive potential may be used for either movement or fire power, the modeler must assign a reasonable amount of potential for a given weapon system. For instance, compare a light infantry unit with a combat helicopter unit. Each unit will be represented as an entity. Assume a maximum limit of 100 offensive potential points are allowed for any entity. In this case, a helicopter unit has high mobility and high firepower, while the light infantry unit is rather slow and armed with light weapons. Assigning a helicopter entity a full 100 offensive potential points, and 10 to the light infantry entity, will give each a reasonable difference in offensive potential or capability. In addition to this, the modeler also has the freedom of assigning how much each terrain type can charge each specific entity type for transit. A helicopter may be charged one point going through a swamp node (since helicopters

actually fly over swamps) while the light infantry entity will get charged 10 points.

Numbers, Predictions and War, by Trevor Depuy, is a good source for relative combat

capabilities of various unit types. (Depuy, 1985: 185-231)

The simulation is set up so that each side will complete a 'turn', then the other side

responds with a turn. This sequence will go for a fixed number of turns. A 'turn' is

made of two parts for each entity. First, the entity executes its movement command, then

if able, executes a combat action against an enemy force. There are two types of combat

action. Every entity has the ability to resist any enemy entity invading its node. There is

no charge against either entity's offensive potential for this node battle engagement. The

modeler can give each entity type a special ability to engage enemy entities at distant

nodes. This is called 'Ranged Fire'. Ranged fire costs the firing entity a fixed amount

offensive potential. Hence, the entity's allocated offensive potential may be used for

movement or ranged fire. Unused potential is lost at the end of the turn.

Ranged Fire. Along with movement ability, ranged fire potential defines the

uniqueness of each entity type. An armor entity as well as an artillery entity can fire

projectiles, each with their specific capabilities. The user specifies the *range, accuracy*,

and the *strength* of the projectile effects to describe each entity type's ranged fire. He

also assigns the offensive potential cost of each shot. Consider the previous helicopter

entity example. A realistic ranged fire capability is the ability to fire guided rockets – a

projectile. The modeler can assign the rockets a range of four node lengths, an accuracy

of 50%, and a strong '80' for projectile effect (1 to 100 relative scale). After a trial run,

the modeler can keep the capabilities or change them to any desired level. Consider the

helicopter entity with 100 offensive potential points. If the cost for firing is relatively low, say one offensive potential point for each shot, the helicopter entity can spend 90 points for movement, and still fire 10 rockets (expecting 5 hits).

Ranged fire can be further modified to consider the vulnerability of the target being fired at and the terrain the where the target is located. Specified accuracy and effects can be modified in consideration of the target's ability to hide and protect itself. For instance a helicopter cannot expect the same results against an armor unit in a forest as out in the open. Accuracy and effects may both be adversely affected in the forest. A simple adjustment for this particular type of engagement can be made without modifying the overall reference settings. *Empire II* gives the user much flexibility in overall definition of an entity type's true offensive firing capability.

Combat Strength. When an entity fires on an enemy entity, the computer will randomly draw to determine if a hit occurred. The result is influenced by the specified accuracy of the fired weapon. Given a hit, the computer also draws randomly to determine the effects of the hit. The target's '*Combat Strength*' is decreased accordingly. The modeler determines the initial combat strength value for each entity type. The combat strength value determines the entity's offensive firepower, morale status, and well being. Again, consider the helicopter entity. The helicopter entity actually represents a combat unit of helicopters. The entity can absorb a few hits, and will lose combat strength with each hit. The entity will remain alive, but with 'fewer' helicopters. The entity can still fight but not at its original strength. If the strength drops below a certain level (as determined by the modeler), an entity can become demoralized and attempt to leave the battle. Complete loss of combat strength equates to a killed entity.

Force attrition is calculated in a way not unlike the ATLAS force attrition model formerly used by the US Army (Hartman, 1985: Chap 2). Here is an example of how the calculations work (New World Computing – User Manual, 1995: 31):

1. The firing entity Ranged Fire Strength is modified by the ratio of its current Combat Strength to its original Maximum Combat Strength.

$$RFS_{Current} = RFS_{Original} \left(\frac{CS_{Current}}{CS_{Original}} \right) \tag{29}$$

2. Once the shot is taken, Ranged Fire Accuracy, as modified by target conditions (weather, target camouflage, etc.) determines if the target entity is hit.

$$P(Hit) = RFA_{Firer}(Conditions_{TGT}) \tag{30}$$

3. Given a hit, the target entity's Combat Strength is modified downward as determined by the firer's Ranged Fire Strength and a corresponding Combat Effectiveness value of the target entity. Each firer is assigned a Combat Effectiveness value against each particular enemy entity type. If this value is high, then a hit will cause great damage to the target.

$$CS_{Current} = CS_{Original} - RFS_{Firer}(CEV_{Firer/TGT}) \tag{31}$$

If the target's combat strength is brought down below the kill level, the attacker scores.

Definitions. The modeler may also define some basic characteristics for the opposing sides. *Efficiency* reflects the training quality of each side, and a single value will apply to all entity types of a given side. *Combat Effectiveness* defines an overall reference value

for node engagements (not ranged fire) for all entities on the battlefield. It gives the computer a reference to calculate damage between entities engaged in a node battle. All damage calculated is assessed against the respective entity's combat strength value.

Each entity type is allowed an option to entrench itself. The entity becomes immobile but is still free to use ranged fire. An *entrenchment* value is defined for the entire battlefield so that the computer can factor in an attacked entity's dug in defensive posture on any combat strength calculation. On the other hand, selected entities can go into a *transit* mode that accelerates their movement but takes away any ability of ranged fire. If one of these units gets attacked, the computer references the transit modifier in the calculation. In this case the effect of the hit will be enhanced!

Detection and *identification* are also considered in this simulation, although line of sight calculations are not used. Instead, reference detection and identification distances are defined for the battlefield. Each entity type's detection ability is then modified based on that reference. Factors that affect the ability to detect and identify include terrain type, weather conditions, and the type of entity to be detected. The helicopter entity can be allowed to have great detection ability for an armor unit in the open, but restricted ability against a light infantry unit within a forest node. Restrictions can be enhanced for unfavorable weather conditions.

Supply and Reinforcements. *Empire II* features added realism by considering logistics and reinforcements. These features are useful in longer running simulations. In the models considered in this research, units were considered fully supplied at all times with no reinforcements available.

Measures of Effectiveness. *Empire II* scores the battle in a combination of three ways. Points are awarded for killing enemy entities, capturing land mass, and capturing strategic locations. The user can use what ever combination that he desires to score the battle. Point values are awarded for each entity killed and each strategic location held. Possession of a strategic location can be scored once in each game or with every turn. Any one of the three groups can have a multiplier to weight the final score. The modeler defines how the scoring takes place consistent with the MOE desired.

Conclusion

This chapter discussed the RSM modeling technique and how it can be applied to map a simulation model. This mapping can be used for a number of applications. In this research, RSM will be used to find the sensitivity of a combat simulation outcome to numerous input factors. Instead of using an established combat simulation, *Empire II*, a commercial wargame is utilized. The computer wargame turns into a simple combat simulation by letting the computer handle all the human interaction. *Empire II* allows the player to custom build any tactical level scenario. RSM can be used to find what inputs the simulation is sensitive to. The technique may also be useful in larger and more complex settings.

III. Literature Review

Introduction

For the last 20 years, computer simulations have been used extensively in the military world for decision support. Simulations have had a great following because of their ability to generate accurate data for a given scenario without committing resources to actually building the scenario, assuming it can be built at all. Simulation models are used to compare alternative inputs, giving the decision maker additional insight of the ramifications of the decision itself. Speculation about possible outcomes are either confirmed or discredited by those simulation models.

But, as simulation models advance and become more complex, blind trust in their results increases. The simulation's ability to generate accurate data is based on the fact that the input data is accurate. Throughout the literature, a question of validation repeatedly comes up. Validation is simply the confirmation that the model is indeed a credible source of information for a particular application. Apparently many of the simulation models, especially those used in the military, have a reputation among simulation and modeling experts of lacking adequate validation.

Validation in itself is made up of several elements. However, no single methodology has been established. In a joint thesis by Capt. Craig Ghelber and Capt. Charles Haley, a validation methodology is proposed which is broken down into four phases (Ghelber and Haley, 1980: 40-60). First, the conceptual phase where the actual application of the model is defined. Once the simulation is coded, a verification phase assures the simulation behaves correctly. The third phase examines the credibility of the

41

model. This includes a sanity check by somebody familiar with the real world system, and a sensitivity analysis to examine the limits of the model. All this accomplished, the fourth and final phase, confidence building. The successful use of the model, once it has been through the first three phases, will increase the users confidence. All put together, in time, the simulation model is recognized as a sound source of information. However, users get so confident of the model that the credibility phase (which should be an ongoing phase) gets neglected.

The Need for Sensitivity Analysis in Combat Simulation

As the computer simulation model ages, the behavior of the model may not reflect current conditions. The reliability range of the model shrinks. In their book The Wargame, Garry Brewer and Martin Shubik indicate there are many cases where models are not maintained at an acceptable level (Brewer and Shubik, 1979: Ch 13). Even when models are developed, questionable validation taints their credibility. Sensitivity analysis is neglected. They say:

> Sensitivity analysis is an important operational control, particularly
> when the number of variables in a MSG (Models, Simulations, and
> Games) is large and the model is complex. It is necessary to know the
> effects of alterations in input parameters on MSG behavior. This is
> especially true when the precision of the data is questionable, where the
> data variance is large, or where random perturbations are treated
> explicitly. Nevertheless, 45% of the MSG's we surveyed had not been
> tested for sensitivity. Such testing is generally expensive and time-
> consuming, but it is essential to a determination of the MSG's
> usefulness and validity. (Brewer and Shubik, 1979: Ch 13)

The editors of The Military Applications of Models responded to some similar comments by Brewer and Shubik. They commented that some work has been

done, but "institutional arrangements are not conducive to generation and funding of such a program" (Battilega and Grange, 1984: 521).

These comments were made 10 years ago, so clearly with lack of funding, problems still exist. Sensitivity analysis is one of many ingredients in the validation process. It can highlight shortcomings of a model. In addition, it can point out the factors that need to be closely monitored. Methods of sensitivity analysis tend to be cumbersome. Several techniques are available, each with their own advantages and disadvantages. One sensitivity analysis technique uses Response Surface Methodology. RSM has proven successful in the following examples.

Previous Efforts in Sensitivity Analysis of Simulations

Ronald Iman, Jon Helton, and James Campbell of Sandia Labs uses a response surface technique to analyze the sensitivity of a computer simulation output (Iman, Helton, and Campbell, 1981: 174-183, 232-240). They indicate the benefits of such an analysis include:

1) An indication whether the model operates as intended
2) Identification of unimportant variables or unnecessary model complexity
3) An assessment of relative input variable importance for guidance in data collected

Dr. Iman points out the term 'sensitivity analysis' is widely accepted as meaning the study of variation in model output with respect to input. Overall, he wanted to answer the question – "What variables have a significant effect on model output and what is their relative importance?"

Data from a computer simulation model was used to generate a response surface metamodel. A metamodel is a model of a model. In this case, the metamodel was used to map out the responses of the simulation for various inputs. A stepwise regression process was used to determine the coefficients of the response surface's function. The coefficients indicate the sensitivity of the simulation's outcome to their respective factors. Once the metamodel was confirmed to adequately map the simulation's behavior, the most significant factors were simple to see.

The article concludes with a generic 'checklist'. This will be the general procedure followed in this thesis. Although the article specifies that this is not an algorithm for all sensitivity analysis situations, the checklist will give a good basic chronological outline. The checklist is broken into five steps:

1. Define the objective of the sensitivity analysis, then identify which input variables are to be considered, and the desired simulation response to measure.
2. Select input vectors for the regression equation from some experimental design.
3. Perform a preliminary factor assessment – find which factors are important.
4. Determine relative factor importance by constructing a response surface through some regression technique.
5. Check the predictive ability of the resulting response surface metamodel against actual simulation model results.

Alan Goldfarb used this general outline in building a response surface for a discrete simulation model (Goldfarb and others, 1988: 689-693). The problem was that certain factors in a production process significantly affected the production rate, but he did not know which ones. He had narrowed a production rate problem down to 22 parameters

for consideration. Using a typical full factorial sampling technique, the required runs of the simulation were excessive (over 4 million runs!). Instead, a Monte Carlo method of setting factor values was used.

Each of the 22 factors could be described with a triangular distribution. Using random numbers, he selected a value for each factor in a given simulation run. Note that this particular simulation model used is a deterministic model. The model will calculate the same outcome every time for a fixed set of factor values. If any of the values are changed, a different outcome is calculated. 50 separate deterministic runs, each with their unique set of values for the 22 factors, were completed.

A stepwise regression technique was used to determine that seven of the 22 factors explained 85% of the production rate. In other words, the production rate (the response) was most sensitive to those seven factors. Therefore, management could now be alerted to focus on those seven factors.

But, before the results could be presented, a check of the response surface was required. The author subjected the regression model to the same 50 sets of parameter values, but limited the regression function to just the seven factors. The metamodel was found to adequately predict simulation results – thus giving the response surface credibility as a good model of the simulation.

Probably a better way of testing the regression equation in this case, is to make another set of runs using other randomly selected settings for the 22 parameters. These values would be completely independent of the values used to make the metamodel in the first place. Then compare the seven factor regression model against those simulation

results. The cost of this would have been more runs over the original 50 on the computer.

Other examples of RSM use in sensitivity analysis include the Ghelber and Haley thesis. As mentioned before, sensitivity analysis was an important factor in the credibility stage of their validation methodology. A response surface was constructed to determine the main factors of a stochastic air defense model. It was confirmed that this step was very time consuming and was recommended that an additional look at the process would be beneficial (Ghelber and Haley, 1980: 61).

The Myers thesis used RSM sensitivity analysis as an aid to a deterministic decision model. Other sensitivity tools were used in the research and found the RSM method highly successful. The method was used in lieu of individual factor sensitivity analysis. The major benefit in this particular case, was an accurate regression model derived from a minimum of simulation trials (Myers, 1995: 5-2).

A parts quality thesis by Matthew Stone used a three factor RSM model to isolate which factor the outcome was most sensitive to. He concluded in the end that two of the three were very significant, but he also made an interesting observation. With only three factors, a full factorial design is quite feasible (Stone, 1988: Ch 4). This gives an advantage where all interaction combinations could be considered for their effects on the outcome. This could be a significant direction to look. However, Dr. G. Taguchi, a renowned name in the experimental design field, stresses the point that individual factors account for most variability in a response. An experimenter should first isolate the main factor effects before embarking into larger experiment designs aimed at finding interaction effects (Taguchi, 1987: Introduction).

RSM and Sensitivity Analysis of a Combat Simulation

Two recent thesis efforts use RSM techniques to build metamodels of separate

THUNDER combat simulations. Capt. Ryan Farmer concluded in his thesis that a

Resolution V experiment is satisfactory in producing a metamodel (Farmer, 1995: 6-3).

However, he learned large ranges between the high and low experiment values can create

a problem of an excessive design space. A lack of data in the voids between the high and

low values of a two-level experiment leaves a note of uncertainty about the actual

metamodel surface. Are there significant simulation model characteristics between the

two-level design points left unexamined? In Capt. Farmer's case, the large design space

did not hinder the results significantly. The design space for his research used each

factor's reference value plus and minus 40%. This research will examine a space defined

by the reference values plus and minus 10%.

Although the design space may not give the overall picture, a simulation model's

sensitivity can be accurately demonstrated over a limited design space. A linear model

may be examined over a large design space or a very small space. Either way, the

response's sensitivity is the same. In a combat simulation, relationships between input

factors and the response may be linear. If the experiment is referenced near the nominal

levels of each factor, then a small ($\pm10\%$) design space can produce a linear metamodel

that will closely approach the actual response surface regardless of its linearity. The

factor coefficients can then be interpreted as the response's sensitivity to the various

inputs. With a linearity assumption, sensitivity may be extrapolated over the entire

combat model. However, if the combat model behaves in a non-linear fashion then the

metamodel is usable only in the design space. Follow-on experiments will be required to provide adequate data to form an accurate non-linear metamodel.

A second thesis effort by Capt. Steven Forsythe provides a methodology that not only provides a metamodel of a combat simulation but also optimizes the response (Forsythe,1994: Ch 3). The response calculated was rather linear although an interaction term did have a 'significant' effect. The slight nonlinear nature of the metamodel allowed an optimization attempt. The result is somewhat misleading. The optimized response was well outside the design region. This information identified a point outside the feasible region as a point of diminishing returns. Operationally, this information is rather useless for two reasons. First, the metamodel, which is nearly linear, indicates that with more weapons employed the better the results – right up to the constraint. A linear model would provide the same information. Second, even if there was a optimal result in the feasible region, it is accurate only for the fixed enemy force provided in the combat simulation. Consistently using the same 'optimum' strategy against a smart enemy would certainly yield non-optimal results. Claiming an optimal strategy based on a fixed enemy is unwise!

Putting this in layman's terms, imagine a hockey team. This analysis is similar to analyzing the optimum number of hockey players on the ice at one time. With data recorded for four, five, and six players on the ice, a slightly non-linear response surface is generated for predicting the amount of time between goals. Calculations could show a result where 15 players on the ice would yield best results, even though only six are allowed. Common sense indicates, the maximum number of six players allowed on the ice is certainly the best strategy. Like hockey, the optimum combat strategy will likely be

found at the constraints – regardless of the enemy. Precise predictions of a nearly linear and highly variable response are misleading. A nearly linear response surface based on such data may as well be treated as linear.

The Relationship of Combat Simulations and Commercial Wargames

Like weather simulations, combat simulations are based on a high degree of uncertainty. Claiming a set of inputs will result in a certain response carries a significant degree of risk. Warriors however have had a history of using simulations to identify and improve their advantages. In his most recent wargames handbook, James Dunnigan, a noted wargame advocate, states that combat simulations and wargames are major tools for the Pentagon and the warfighting CINC's (Dunnigan, 1992: 234-264).

Mr. Dunnigan states that wargames are merely a 'playable simulation', while a combat simulation "is a model, or a collection of models that can be easily manipulated to test the 'what if' questions". He goes on to explain that combat is very unpredictable, but like weather simulations, a moderately high degree of confidence can be afforded to their results. But, he points out the military strive for the unobtainable perfect prediction model is infeasible. "Until recently, professional wargaming tended to go after the less than 1 percent deviation in accuracy and consume years and millions of dollars in the process." He further explains, "Commercial manual wargames can generally achieve a '5 percent solution'", while a wargame on a PC can get a "nearly 1 percent solution". In other words, a commercial PC wargame is an extremely economical combat simulation that can give legitimate results.

This research does not intend to use the commercial PC wargame *Empire II* as a predictor of combat engagements, but rather as a vehicle to show how RSM may be used in conjunction with a combat simulation. However, if the game results appear to be consistent with other military combat simulations and actual experience, then consideration should be given to Mr. Dunnigan's thoughts.

Conclusion

A major benefit of a simulation is the ability to demonstrate sensitivity to various inputs. The sensitivity of the model must be consistent with actual process behavior before a simulation model can be declared valid. But, simulation models themselves must be analyzed periodically to confirm their currency with the actual process. A valid model can explore regions where the actual process has not been. A combat simulation can be used to determine the effects of changing force structure as well as weapon characteristics to a specific response. Mapping the response can give a quick turn analysis without referring to the simulation model repeatedly. However, the metamodel produces only a prediction of a simulation trend. Claiming an optimal solution to a highly stochastic combat simulation (which in itself is a model of a highly unpredictable process) is subject to criticism. Metamodels of highly stochastic combat simulations cannot be used to confidently predict short term results, but can isolate trends influenced by a limited number of factors.

IV. Methodology

Introduction

The goal of this research is to define a methodology that identifies weapon systems and their characteristics which have the most influence on a simulation outcome. A good metamodel produced by this methodology will adequately describe the behavior of the simulation model based only on those most significant factors. This chapter will describe the methodology from the building of a simulation scenario to the actual mapping of the simulation behavior. First, an overview of the simulation scenario is covered. All weapon systems and their characteristics are defined. Each individual weapon characteristic is treated as a separate factor. A screening experiment is performed and data collected which will isolate those few factors that are indeed significant to the simulation outcome. To find the extent of their significance, those remaining factors are further analyzed with a follow-on experiment. With these results, a regression metamodel, which maps the behavior of the combat simulation model, will be calculated. Since the regression model is based on a small number of representative factors, a validation experiment is required to determine if the simple metamodel is accurate enough to be useful. If the metamodel is indeed valid, then conclusions can be drawn about the simulation's behavior based only on those most significant factors. Each step of the process will be discussed in this chapter.

Scenario Setup

This section explains the actual scenario used for the research. The scenario defines the battle scene to be simulated. Each weapon type is briefly described in the technology overview. The Battlefield section describes the geography of the area and force setup. Next, the Measure of Effectiveness section will define the scoring of the simulation outcome. Finally, the Attack Scenario section will focus on the actual attack force structure of the scenario.

The Scenario. A generic battle scenario was required to represent not just a real world possibility, but also a scenario which is not uncommon. Several scenarios were considered. Consider the possible objectives of two sides. Generally, each side wants to control the same objective. The objective in this scenario is a strategic location.

The first scenario put either side facing each other with an objective in between. The goal is identical for both sides – capture the strategic location before the other. This turned out to be rather unrealistic. The only situation that resembles this is the big westbound push by the Soviets, and eastbound push by the other allies, to capture as much of Nazi Germany as possible before meeting each other. This did not degenerate into a ground battle between the allies when they met.

A second scenario is to assign both sides each a strategic location to protect, and the desire to capture the other. However, a dual objective is rather unrealistic. Both sides are playing the aggressor role, although there is an objective just as precious to defend. This condition is common in various sports such as soccer or hockey, but not in conflicts.

52

The scenario selected involves one aggressor force (Blue) whose mission is to capture a strategic location protected by a defensive force (Red). This circumstance is rather common in world conflicts. In Desert Storm, after the Iraqis played the aggressor against Kuwait, the coalition then played the aggressor against Iraq – two conflicts separated by months of stalemate.

Operational or theater level simulations such as TACWAR and THUNDER have 'piston driven' structures where the campaign is separated into individual corridors, each independent of the other. The forward line (known as the FLOT) moves like a piston in each corridor subject to each individual battle. The overall result is calculated by aggregating all the corridor FLOT positions into a single 'continuous' (but jagged) FLOT position. *Empire II* offers a simulation tool at the tactical level or the level where the battles are fought. Consider the conflict in this simulation as a battle in one of the individual corridors.

Technology Overview. Each side is made up of individual unit types of identical strength. The main difference between the two sides are the numbers of units. This is a summary of each unit type and their capabilities.

1. *Light Infantry* – Dismounted troops armed with light firepower. Dismounted troops are used by the defenders to protect urban areas from the aggressors. Red Light Infantry will force the Blue aggressors into urban warfare if Blue penetrates that far. Since they can disperse and defend throughout the town, they are considered somewhat survivable. Aggressors also have light infantry but are considered mounted until they arrive in the town itself.

53

2. *Mechanized Infantry* – Mounted troops on board fighting vehicles armed with weapons that have medium range (4km), moderate fire rate, and moderate fire power. Nearly half of each force is composed of mechanized infantry. Mechanized infantry is very mobile and is not considered a soft target, so artillery will have a hard time killing them with their area fire. However, armor and helicopters can engage directly with good success. Mechanized infantry possess anti-tank weapons and some anti-air missiles, posing a formidable threat to their adversaries.

3. *Armor* – Compared to mechanized infantry, tanks are armed with lower range (3km) weapons, but with strong fire power and a slightly higher fire rate. Like mechanized infantry, armor makes up nearly half of each ground force. Tanks are capable of moving as fast as mechanized infantry. Coupled with heavy protection, armor is a tough target to kill. Aviation units, fighter jets, and other armor units are considered as major threats, although any unit can possess at least some weaponry capable of putting tanks out of commission.

4. *Artillery* – Long range guns (13 - 17 km in this simulation) with ballistic projectiles. Artillery batteries are allocated to each brigade. Each artillery unit is allowed numerous shots, but since they fire into an area target, accuracy is not considered high. However, given a hit, a soft target will be damaged heavily while an armored target will be somewhat hindered. These artillery units are minimally capable of moving while firing, but can be put into a transport mode to allow them rapid repositioning if required. These units are vulnerable, especially when stationary and in the open, but are capable of limited self defense against all other units.

5. *Aviation* – Consists of various helicopters, a highly agile weapon system that also carries a powerful punch. Aviation units are capable of firing guided missiles more destructive than artillery shells and with improved accuracy. Helicopters can go behind enemy lines and engage targets while maintaining some standoff capability. But, unless they elude well, helicopters are very vulnerable to anti-aircraft fire. Like artillery, aviation is a limited resource and is an allocated asset.

6. *Close Air Support (CAS)* – In addition to helicopters, the Blue aggressors are given the added advantage of close air support. CAS in this simulation consists of jet fighter attacks on units at and behind the front line. These equate to very long range artillery pieces that are directed against point targets. Because of the coordination required between the ground commander and pilot, not to mention the CAS skills of the pilot, accuracy is somewhat limited. Given a hit, any ground unit can be severely affected. Close air support can also be used by the aggressors in this simulation to soften defenses well behind the FLOT in preparation for invasion.

7. *Command and Control (C^2)* – The defender's strategic location. C^2 is simulated as a bunker location that is well fortified and defended with short range fire. The C^2 process in itself is not directly simulated.

The Battlefield. The selected scenario could occur anywhere in the world, so rather than simulating at a particular world location, a generic location was developed. *Empire II* offers an opportunity for the modeler to build a generic battlefield location. This scenario is placed in a location where the objective town is found on a seacoast (Figure 15). This town has a strategic location (consider it a command bunker) which is

protected with a Red division made up of two brigades each of armored vehicles and mechanized infantry. Located in the town are several Red battalions of entrenched light infantry. Moving west out of town and away from the seacoast is a line of hilly terrain covered with forest and meadows. The hills are approximately 10 km inland and extend from north to south paralleling the seacoast. Moving further west another 10 km, the hills give way into forested plains that are spotted with lakes and occasional swampy areas. Two divisions of Blue aggressors, each separated into three brigades are located 10 - 20 km west of the hilly terrain. Several passes through the terrain are available to the aggressors although the hills are not impassable.
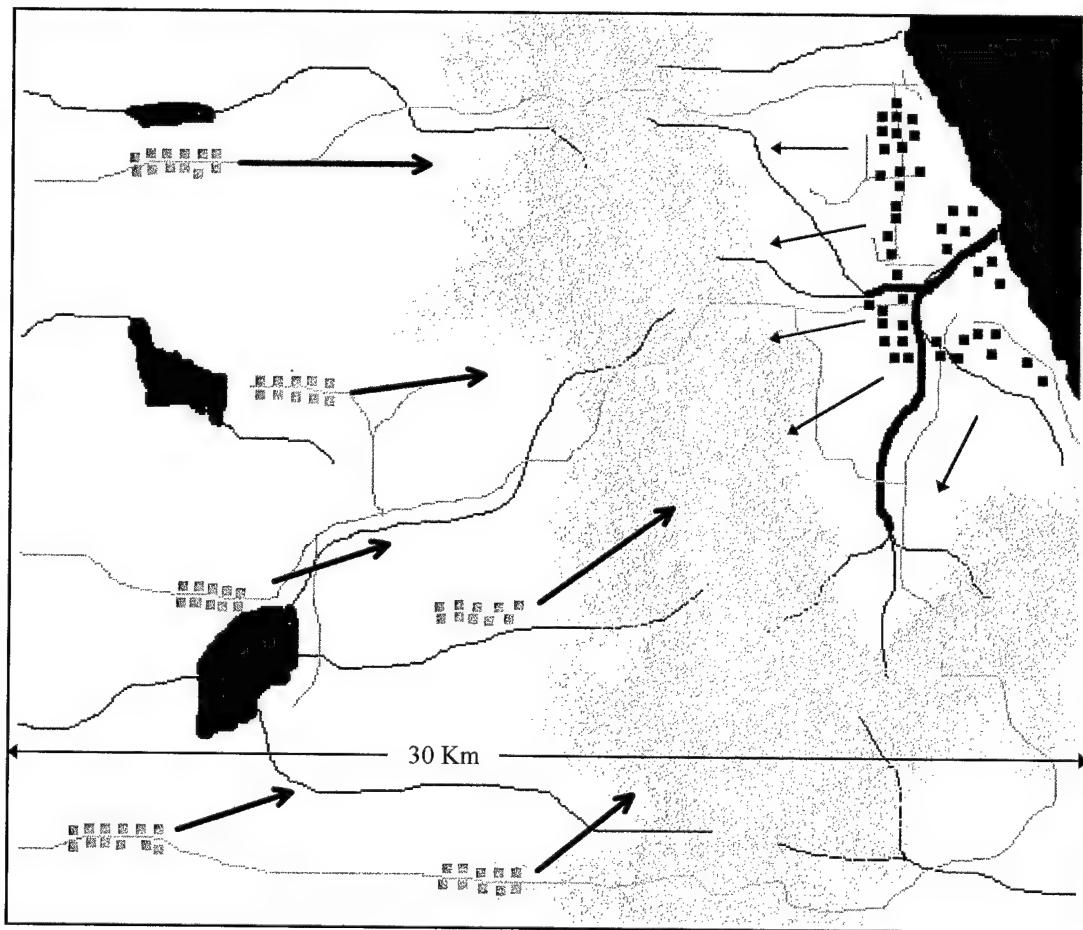
Fig. 15 Battlefield Setup

Measures of Effectiveness. The only way to score in this wargame is to outright kill

an opposing unit entity. Although a shot can hit and disable an opponent, only a kill will

score. Each unit type has been assigned a value based on its offensive potential. This

potential is based on its offensive capability or movement value, its combat resilience or

strength value, and its ranged fire strength. Refer to Table 1. The product of the three

terms are recorded in the M*S*R column. To align the values to a scale of 1 to 25 (the

simulation's scoring limits), two adjustments are made. First, to reduce the magnitude of

the differences but still keep the relative values intact, the square root of each value is

taken. The result is then adjusted down to usable value by dividing by an arbitrary

constant – 25. The result was then truncated up to the next highest integer.

Table 1 – Scoring Value Worksheet

	Movement Value	Strength Value	Ranged Fire Value	M*S*R	Square Root	Sqrt/25	*Score Value*
Light Inf	2	50	2	200	14.1	0.6	*1*
Artillery	20	30	20	12000	109.5	4.4	*5*
Mech Inf	15	40	15	9000	94.9	3.8	*4*
Armor	15	70	25	26250	162.0	6.5	*7*
Aviation	60	65	35	136500	369.5	14.8	*15*

Since Blue's Close Air Support airfields are out of Red's range, no value is required.

The Red defender's Command and Control is the Blue aggressor's objective, thus is

given a maximum value of 25. The simulation terminates after a set period. Final scores

are summed for each side and presented as offensive totals for Blue and Red.

The Attack Scenario. Several attack scenarios were considered in this research. Typically, Army wisdom insists on at least a 3-1 advantage before an aggressor engages. This creates a problem for this research. The simulation outcome tends to become more robust as advantage increases. The outcome's sensitivity to various factors becomes impossible to detect. In battle, this is very important point for the aggressors. The bigger the advantage, the many various factors become less significant to the final outcome – thus insuring victory.

On the other hand, a slight advantage to the aggressor yields victories in only half the battles. Even slight changes to the factors show significant but inconsistent differences to the outcome. For this research, an equal strength battle seemed to be a logical way to go. However, this lack of realism where an aggressor opts to attack with no advantage would put any research results in question.

An interesting situation occurs in the simulation when the two sides are nearly equal. Even though victories are evenly split, the margin of victory tends to be rather large regardless of who was victorious. These results tend to follow Lanchester-type attrition models. The Lanchester-type attrition model states the attrition of one side depends on the strength of the other side. As one side gets an advantage, that advantage accelerates until the other side is vanquished. If neither side gets an advantage, the battle outcome will pivot on the last two standing. But in these simulations, sooner or later one side gets an advantage which always increases through the final turn. This presented a problem where results had huge variance. Just by changing a few minor factors, the battle outcome is completely altered. Therefore, as the simulation points out, a commander

may be very unwise to engage a somewhat evenly matched opponent, since the outcome is very sensitive to factors that nobody has control. Attacking with a large advantage negates the problems associated with the highly stochastic fog of war.

With the advantage firmly in the aggressors hands, these factors will still make a difference, but in a much more subtle manner. No doubt the aggressors will be victorious most of the time, but a change in the input factors will result in noticeable differences in the scoring output.

A compromise between the two force ratios is used in this research. An approximate 2-1 advantage is given to the aggressors (Table 2). Although unlikely and maybe unwise for an army to attack with this small of an advantage, it is not unreasonable. The defender's objective is not necessarily to defeat the aggressors, but to hold them off until reinforcements arrive. This scenario is set for a twelve hour day. Twelve hours is the basic time period for a TACWAR simulation. Again, this battle represents only a part of a larger campaign. The battle represents only one of the 'pistons' in a simulation like TACWAR or THUNDER. The 'piston' in this particular corridor is expected to move in a eastbound direction as the aggressors invade the defender's town.

Table 2 – Total Forces in Scenario

	Aggressor	Total Value	Defender	Total Value
Light Inf	0	0	8	8
Artillery	12	60	8	40
Mech Inf	28	112	14	56
Armor	24	168	14	98
Aviation	4	60	2	30
Sums	68	400	46	232

Design of Experiments

The objective of the entire series of experiments is to identify which factors are significant to the simulation outcome, measure their actual effects, and then to validate the findings. Since simulation runs are tedious, an economical design of experiment is a major concern. First, isolating the significant factors is accomplished with a screening experiment. A Plackett-Burman experimental design is very useful to screen large numbers of factors. After identifying the most significant variables, a Resolution V factor analysis experiment will determine the effects of each remaining factor and their interactions. With these two main steps, the factors which the simulation outcome are most sensitive to will be isolated and evaluated. The final results must be cross-checked against an independent experiment to validate that the metamodel is indeed accurate enough to be usable.

Screening Strategy. The Plackett-Burman design is a classic experimental design for screening large number of factors. Some of the design limits include the fact that it is only a Resolution III design and that it has a complex alias structure. Significant factors will stand out, although there is a small possibility that some factor will falsely stand out because of some multi-factor interaction alias. The assumption is made that all interactions are insignificant and only main factors account for different simulation outcomes. In this research, 23 variables are analyzed with a Plackett-Burman screening experiment using 24 trials. The experiment design will be the basis for 'X'.

By using a stochastic simulation, each trial will require multiple runs. Enough runs are required to define an 'average' result for each trial. The Central Limit Theorem states

as the number of trials approach infinity, the sampling will result in a normal distribution. Usually, 30 or more samplings will closely approach a normal distribution (Mendenhall, 1989: 317-320). Fortunately, each factor is set to one of two levels, high and low. Since each factor is considered in each of 24 trials (12 at a high level, 12 low), then three runs for each trial will produce 72 data points for each factor – 36 for the factor's high setting and 36 low. With a highly variable simulation model, even 30 data points may not be enough to adequately define the variance. The variance, and in turn standard error, must be at a usable level before any statistical claims can be made. An 'n' number of runs are required to develop a standard error small enough for the defined confidence level of the 't' test later used for factor selection (Law and Kelton, 1991: 287-290). However, 'n' is an unknown number, so like above, a safe move is to gather approximately 30 data points for each of the factor settings to begin with.

Once the information is collected from the experiment, the **y** vector is at hand. A regression model may now be constructed. Once the model is constructed, a test for residual normality must be performed to confirm that all the statistical tests mentioned in Chapter 2 can be used. The run average for each trial will produce a regression model for all 23 factors based on 24 trials. With only one degree of freedom , the model may give good insight to the most significant factors (despite an R^2 of 1.0). On the other hand, using all 72 runs will produce a regression model that may not have an R^2 of 1.0 but with the additional degrees of freedom, will better screen for those significant factors.

The 23 Factors Considered. Most theses authors before experimented with the number of combatants in a scenario to isolate the effects of changes to the force

61

structure. This thesis concentrates on the effects of altering various weapon system characteristics to the scenario outcome.

In this research, only the aggressor (Blue) weapon systems are altered. Five aggressor weapon systems are considered – Artillery, Armor, Aviation, Mechanized Infantry, and Close Air Support. Each weapon system is represented as an entity type in the simulation. Each entity type, except for CAS, is affected by five characteristics. The first two characteristics are entity movement (offensive capability), and combat strength (the ability to survive). The other three characteristics are Ranged Fire capabilities. Ranged Fire accuracy, strength, and projectile range are all entity attributes considered. Four weapon system types times five characteristics produce the first 20 variables. A CAS entity is actually a stationary airfield that is out of range of the Red forces, therefore movement and combat strength will remain constant. Ranged Fire projectiles are represented by the actual sorties flown. Since the jets have a range greater than the battlefield area, Ranged Fire projectile range remains large and constant. So, only two of the Ranged Fire variables remain for CAS – accuracy and strength. The total list stands at 22 variables.

The last variable considered is Blue Combat Efficiency, or the relative strength of the Blue army compared to Red. Equal sized armies from different parts of the world are not necessarily of equal combat value. This is a rather subjective factor in which the modeler assigns some objective values. In this experiment, the factor is a reflection of the Blue side's training and reliability, which in turn affects their ability to employ. If the outcome is sensitive to this factor, then a closer look should be taken at how the

simulation objectively measures and uses the 'training' or 'functional reliability' value of an army. A full breakdown of the factors is shown in Table 3.

In the experiment, each factor is adjusted 10% higher and lower about the reference value. A ±10% increment is enough to allow a satisfactory look at the sensitivity of all factors despite large simulation variability. If a larger positive increment is used, the response will be constrained by the lack of Red forces. In other words, any high setting will give the same response as all high settings – the entire Red force would always be destroyed, resulting in a constant outcome.

Table 3 -- The 23 Factors and Their Settings

Plackett - Burman Design (N = 23)			-10% Low -1	Medium 0	+10% High 1
	Factor				
Artillery	1	Movement	9	10	11
	2	Combat Strength	22	25	28
	3	RF Strength	22	25	28
	4	RF Range	13	15	17
	5	RF Accuracy	22	25	28
Mech Inf	6	Movement	13	15	17
	7	Combat Strength	36	40	44
	8	RF Strength	13	15	17
25%	9	RF Range	3 (-2.5)	4 (0)	5 (+2.5)
	10	RF Accuracy	54	60	66
Armor	11	Movement	13	15	17
	12	Combat Strength	63	70	77
	13	RF Strength	22	25	28
30%	14	RF Range	2 (-3)	3 (0)	4 (+3)
	15	RF Accuracy	63	70	77
Aviation	16	Movement	54	60	66
	17	Combat Strength	58	65	72
	18	RF Strength	31	35	39
	19	RF Range	13	15	17
	20	RF Accuracy	63	70	77
CAS	21	RF Accuracy	45	50	55
	22	RF Strength	27	30	33
Efficiency	23	Aggressor Efficiency	-1	0	1

Two factors could not be incremented by only 10%. Since the simulation model will only allow discrete integer values for input, relatively small values such as the Ranged Fire Range for armor and mechanized infantry could only be incremented by 30% and 25% respectively. There are two ways to compensate. First, adjust the final regression results for each factor down accordingly. For instance, when the factor is three times the standard increment, divide the resultant coefficient by three. A second option, used in this experiment, adjusts the **X** matrix to reflect the non-standard increments. Instead of having all (-1, 0, 1) in the matrix, the column that has three times the standard increment will have a series of (-3, 0, 3) in its column.

Analysis of Factors. Once the most significant factors are identified and ranked, a certain number of them are chosen for further analysis. Analysis of those remaining factors will create the final mapping that will ultimately describe the sensitivity of the simulation's outcome. A subjective decision based on either a significance criteria (α), or an experiment design limit (such as a budget and/or time constraint) will determine which of the screened factors are further considered. Once the most significant factors have been isolated, their interaction effects should also be considered for their possibility of significance. Selecting a number of finalist factors to fit at least a Resolution V experiment is preferable. With a minimum of simulation runs, this experimental design will allow all two-factor interactions an opportunity to show themselves as significant without being aliased with a main factor.

Using 95% significance criteria based on the screening experiment's 't' test, a certain number of factors are identified for further consideration. Generally, only a small

fraction of the original main factors will meet the selection criteria. With luck, the number of factors selected will fit nicely into a Resolution V experiment design. If not, the most significant of the factors which do fit a Resolution V experiment should be identified. A possibility exists that there are a shortage of significant factors to fill the experiment's design. This is an opportunity, with no extra cost, to add factors that don't quite meet selection criteria. In other words, if a 16 trial Resolution V experiment can handle five factors, yet only four pass the screening criteria, a fifth factor can be added at the experimenter's discretion.

Again several runs have to be made for each trial. In the case of five factors in a 16 trial experiment, at least four runs should be made for each trial, a total of 64 runs. That results in 32 data points for each of the two factor levels. Center trials are added for a lack of fit test. Three to five center trials are recommended (Myers and Montgomery, 1995: 311). Five trials with five runs each will result in 25 data points for each of the factor's reference or median value.

With the Resolution V experiment designed and completed, both the **X** matrix and **y** vector are identified, and an initial regression analysis can be accomplished. Like the screening experiment, the residuals must be normally distributed before the adequacy tests of Chapter 2 can be used. Assuming the residuals are normally distributed, the tests can give a measure of how good the regression model is. Unlike the Plackett-Burman screening experiment, a limited number of factors coupled with numerous trials offers multiple degrees of freedom. Without sacrificing model accuracy, each trial average can now be used to simplify the regression calculation. Compared with models using all trial runs, aggregate error will be lower, so the R^2 values will be much closer to 1.0.

However, average error will be roughly the same, and the adequacy tests will remain essentially unchanged.

The next major consideration is the Lack of Fit (LOF) test. The original model supplemented with multiple center runs may show some kind of curvature in the data set. In the case of a combat simulation, there is an absolute ceiling on scoring (only a certain number of kills are available). A general reduction of the slope should be encountered as the maximum score is reached. The same goes for the case of percentages. Consider a percentage measurement that increases at a nearly constant rate through say 50% to 70%, but the rate must decrease as it approaches 100%. To be credible, the regression model may need to reflect this fact. Granted, data from a restricted portion of the model may not show a significant effect of tapering, but the overall tendency should not be ignored and may need to be considered in the final regression equation.

A failure in a LOF test indicates that interactions may be needed to complete the model. Still a first order model may be all that is needed. Some interactions may help the model to the point where it passes the LOF test. However, if the LOF test still fails, then a second order design is required. Axial trials are accomplished to check each factor individually at each of the two levels. Fortunately, axial trials can be supplemented onto the original experiment when required. In a five factor experiment, that means ten additional trials with the corresponding number of runs each. But, an opportunity arises where the experimenter can be economical with the number of axial trials actually required. If the first order model determined some of the selected factors to be insignificant, those factors can be eliminated. Axial trials for those factors should not be needed for the final second order model.

A second order model will give the regression surface extra freedom to curve and a better chance to fit a set of data. Other methods of fitting a good regression curve include logarithmic transformations. This method will allow a regression line to taper off to a horizontal asymptote. This kind of curve is best for describing a process that eventually reaches a limit. A combat simulation would best be described in this context when design space is large, and the outcome is very sensitive a factor's low setting and robust at its high setting. A problem with logarithmic transformations is that they are confusing to work with and to present. This research concerns only a small design space, therefore a second order model is adequate.

Validation.   Once the regression metamodel has been produced, it cannot be taken for granted. Some kind of validation is needed for credibility. There are two parts to consider in the validation process. First, a comparison against an independent experiment has to be performed. This experiment will provide validation data to test the consistency between the regression metamodel and actual simulation results. A number of different methods can show if a metamodel is consistent with the actual simulation model. They can range from a random draw of factor values to an experiment identical to the original. . Second, a common sense inspection is needed to detect any counter-intuitive results. Since counter-intuitive results can in themselves destroy a simulation model's credibility, some effort has to be put in to explain them. Credibility relies on good validation, so enough of a budget must be allocated for a satisfactory cross-check.

67

Once a metamodel of some corresponding simulation model has been constructed, the first step in validation is to demonstrate that both behave in the same way. The regression metamodel is a very simplified mapping of the complex simulation model. But, if the mapping is an accurate representation of the simulation, then the metamodel can then be used in lieu of the simulation model. The metamodel can also be used to spot certain trends of the simulation's behavior – such as sensitivity!

Designing a set of experiments identical to the original set of experiments would be the most thorough way of testing. If a similar metamodel were produced then the regression model is valid. However, that is probably the most expensive way to go. Another option is to take a sampling of random factor settings and compare the metamodel's prediction against actual simulation results. A good cross section of these will adequately compare the two models. A third and most economical way to test is by using the original Plackett-Burman design. Although there are a limited number of total runs, enough are available to draw conclusions. Best of all, the runs are already done! A second advantage is that all of the original factors are considered in the original Plackett-Burman design. If the 'insignificant' factors really are insignificant, then metamodel should be consistent with the simulation! However, a bias can be introduced since the 'insignificant' factors are not randomly adjusted from trial to trial.

Once the metamodel is validated, it may be a dependable indicator of counter-intuitive behavior in the simulation model. Counter-intuitive behavior exhibited in the metamodel may put the simulation model itself in question, even though the simulation

may be quite accurate. The simulation model may indeed be accurately showing something that nobody expected. On the other hand, the actual simulation model may be flawed. Either way, the simulation model must be 'verified', to be sure that it is working correctly.

Verification is a time consuming process of observing the simulation model at work and making repairs where needed. Verification also involves observing the model and finding legitimate reasons why a simulation is producing counter-intuitive results. Either way, a counter-intuitive result is the only real flag that something is wrong, even if nothing is wrong. The modeler must put forth an effort to solve the question.

Whatever validation methods are used, determining if the simulation model and its corresponding regression metamodel are valid is a subjective judgment call. The bottom line questions are:

1. If the simulation's behavior is counter-intuitive, is the simulation model really accurate?

2. Does the RSM regression metamodel, even though not perfect, provide an adequate representation of the simulation's behavior?

Positive answers for both questions will give the simulation model and the resulting metamodel real credibility.

V. Results

Introduction

The research project uses the methodology outlined in Chapter 4. This chapter discusses the results. Each step in the methodology will be quantified so the reader can follow how the conclusions were found. The technological attributes of each entity type used in this particular simulation scenario will be documented. These attributes account for all of the factors used in the research. The actual Plackett-Burman design results which describe how the 'finalist' factors were selected are presented. The findings of the follow-on factor analysis experiment will show how only three of the original factors significantly affect the simulation outcome. Some three-dimensional response surface graphics will visually describe the effects of these three factors. Finally, results are validated. One of the final three factors has a counter-intuitive effect on the simulation model, and is discussed. The regression model coefficients are validated and the predictive ability of the model is tested.

The Simulation

Before starting the presentation of data results, an explanation of the data source is advisable. The experiment involved entity types and their technological attributes. Each entity type started as a generic weapon system with no real power. Each entity type was then defined to closely resemble an associated weapon system. These technological definitions, which made each weapon system unique, are the very factors that were

varied for the research. Appendix A is the technology report for all the entities used on the battlefield. This report is separated into four sections;

The Rules – These are some of the basic battlefield reference values that are simultaneously defined for both sides. These remained constant for all simulation runs. Combat Damage Effectiveness is directly (as modified by the firing entity type) used in determining the amount of damage of a given hit. All modifiers in the report are simply multipliers that allow an overall reference value to be adjusted either upward or downward by the amount indicated (1 = +10% and -2 = -20%). An increase of the entrenchment modifier increases combat resilience while an increase of the transit modifier makes the entity more vulnerable. Activated morale rules allow a unit to run when their combat strength is down to a certain level. Sighting rules dictate an overall reference distance value for both detection and identification (which can be later modified by each entity type).

1. Forces – The only factor considered here is the Combat Efficiency value. This number was adjusted only for the aggressor (Blue) side. A +1 increases the strength and firepower of Blue by 10% across the board. This value accounts for the overall combat ability of a side's army. For instance, the US Army is probably better trained and equipped than an equal sized Arab army.

2. Unit Type Summary – This section gives a technology overview of each entity type. The summary reflects some of the relative strengths of various entities. Some aspects investigated in the research include Speed (Movement value), Maximum Strength (Combat Strength value), Ranged Fire Strength and Accuracy. All other values in this section remained constant.

3. Unit Type Details – Each individual entity type is thoroughly described in this section. All information from the Unit Type Summary is repeated. Nevertheless, more information is presented that will further clarify the capability of each weapon system. The only factor investigated not mentioned so far is each entity's Ranged Fire Range, whose reference value is listed only in this section. The intent of the section, however, is to thoroughly describe the modifiers. These modifiers fine tune those capabilities given to each entity type. For instance, a unit type is assigned a 'movement value' or 'offensive capability' value. The value in noted on top of the each entity's subdivision as 'Speed'. This value is modified by the 'Movement Costs' table. The 'Movement Costs' modifiers account for the different kinds of node terrain – the higher the number, the higher the cost to move through. Similar modifiers are made for combat fire effects. A target in the forest may be tougher to hit and kill than in the open. All modifiers are used for each individual shot calculation, which ultimately affects the final score. Since these modifiers are a subset of the main factor settings, and assumed to individually make little difference, they were left as a constant throughout the research. As certain weapon systems are highlighted as having significant influence on the outcome, those respective modifiers should then be further investigated.

These are all examples of various weapon system characteristics that are essentially untouchable to the user in most established combat simulations like THUNDER. Again, the objective of this research is to see which weapon's system characteristics make a difference to a simulation outcome. Those characteristics the simulation is most

72

sensitive should be closely monitored and maintained in the simulation model. The accuracy and credibility of the simulation model depend on it. However, conclusions based on this particular scenario's technology setup may not be consistent with all other scenarios. Only by doing similar investigations with various scenarios and technology setups will a truly accurate conclusion be drawn.

Screening Results

After all factors were identified, a Plackett-Burman screening design was used to determine potential factors for further investigation. Each factor's significance was based on its influence on various simulation trials. The objective was not to measure which side was victorious most of the time. Truly, victory is based on influencing the enemy leadership to submit to conditions. A single enemy killed, even with large friendly losses, may influence the enemy to compromise. However, an assumption was made that higher enemy losses coupled with lower friendly losses tends to be a commander's goal and therefore was used as the measure. So, the simulation's outcome was measured not by victory, but by the battle's final score – a reflection of total kills.

Two Measures of Effectiveness (MOE) were used. Total *Blue* score and *Percent Blue* score. *Blue* score reflects the number of Red defenders killed. *Red* score was also considered, but two problems arose. First, *Red* scores (the number of Blue killed) were relatively erratic. Second, a low *Red* score coupled with a low *Blue* score indicated the battle was slow moving – not that Red (or Blue for that matter) did worse. In lieu of using a *Red* score, the Blue percentage of the total combined score accounted for both problems. Regardless of the battle speed, *Percent Blue* reflected how well both Blue and

73

Red had done. A second advantage featured was the relative stability of the *Percent Blue* MOE.

Even though a Plackett-Burman design can handle a larger number of variables, the 23 factors most likely to affect simulation outcomes were selected. This was strictly a subjective decision based on opinions of various Army officers. The actual design used is presented in Appendix B. Results of the 24 trials are also presented in Appendix B. All factors (except two mentioned later) were adjusted by approximately ± 10%, and were coded as ± 1. Three runs were made for each trial for a total of 72 runs.

A regression analysis using JMP© PC software, supplied by SAS Institute Inc., was used to isolate a number significant factors to a 95% level of confidence. Confidence tests are considered valid since the residuals for both *Blue* and *Percent Blue* appear normal. Normal Probability Plots are presented in Appendix B.

Five factors appeared significant for both MOE's. However, one of them was Mechanized Infantry Ranged Fire Range. Because of a software limit, this factor could only be adjusted by approximately ± 25% instead of ± 10%. To compensate, it was boosted from a coded value of ± 1 to a value of ± 2.5 in the experimental design. Regardless of their coded values, the standard error was the same for all factors in the model. The 'real' error attributed to a ±10% difference for a factor that was actually adjusted ± 25% is 1/ (2.5) (or 10% / 25%) of the standard error. The parameter found for this factor was relatively small, but with a small 'adjusted' standard error to go along with it, the significance of the '*t*' ratio ended up favorably high. Therefore, the '*t*' test for the regression model provided high confidence that the parameter was not zero. On the other hand, this particular factor had little significance on the simulation outcome. Since

the objective was to find the most significant parameters, this factor was deleted

accordingly. Armor Ranged Fire Range, which with the same software limit could only

be adjusted by approximately ±30%, was boosted up to a round value of ± 3. However

this time, even with the model adjustment, Armor's Ranged Fire Range turned out to be

one of the most important influences on the simulation outcome. With its standard error

reduced by a factor of three, the 't' test result proved extremely favorable. So in this

case, despite the non-standard coding, the factor was shown to be very significant.

All together, four factors stood out as the most significant for both MOE's:

1. Armor Ranged Fire Strength
2. Armor Ranged Fire Range
3. Aviation Combat Strength
4. Aviation Ranged Fire Range

Aviation Combat Strength behaved in a peculiar way. As strength was increased,

Blue's score decreased. Since this factor was among the most significant, it was

investigated further. The behavior created an additional challenge for the overall

validation of the simulation model, which is discussed later in the chapter.

Analysis of Significant Factors

With four 'finalist' factors identified, an additional experiment was needed to

thoroughly investigate their effects on the simulation outcome. Even though interaction

effects were ignored in the screening experiment, the interactions of these most

significant factors may have proven to be somewhat significant. By inspection, the four

factors did not have any obvious three-way of four-way relationships. On the other hand,

three two-factor interactions stood out as possibilities. Each inclusive pair of Armor

factors, Aviation factors, and Ranged Fire Range factors appeared to have the most

obvious chances of having significant two-way interactions. Eliminating any chance of a two-factor interaction alias in the experiment, a minimum of a Resolution V experiment design was required.

With only four factors to consider, a full factorial experiment (which is better than a Resolution V design) would have required 2^4, or 16 trials. An eight trial fractional factorial experiment would only yield an inadequate Resolution IV design. Since a minimum of 16 trials were required anyway, a five factor, 16 trial Resolution V design was utilized. So a fifth factor, although not meeting the initial screening selection criteria, was added. Several semi-significant factors stood out as possibilities for the fifth finalist. The one selected was Aviation Ranged Fire Strength, which appeared to compliment the other four factors by adding yet another likely two-way interaction – the Ranged Fire strength combination. The five 'finalists' with their low, median and high values are listed below in Table 4. Notice that Armor Ranged Fire Range continued to be incremented ± 30%, and the others ± 10%.

Table 4 -- The Five Factors for Final Analysis

Factor		Low	Median	High
	Armor			
a	RF Strength	22	25	28
b	RF Range	2	3	4
	Aviation			
c	Combat Strength	58	65	72
d	RF Strength	31	35	39
e	RF Range	13	15	17

Since both MOE's had a capped maximum value, the final models were likely to have a somewhat decreasing non-linear effect. Five center trials were supplemented onto the 16 trial experiment to allow for a Lack-of-Fit tests for various combinations of factors.

Regression Results

Again, JMP© PC software was used for regression analysis. Each MOE, *Blue* and *Percent Blue*, were considered individually. The initial models considered only the five main factors without interaction. Armor Ranged Fire Strength ('*a*') and Aviation Ranged Fire Strength ('*d*') were individually insignificant to both models, and were eliminated from further consideration (refer to Figure 16). Although not quite meeting selection criteria, Aviation Combat Strength still appeared somewhat significant to the *Percent Blue* outcome. Therefore, Aviation Combat Strength was further considered.

Parameter Estimates	*Blue*			
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	160.22857	1.691407	94.73	<.0001
a–Arm RF S	1.725	1.93775	0.89	0.3755
b–Arm RF R	9.8583333	0.645917	15.26	<.0001
c–Av CS	-2.55	1.93775	-1.32	0.1912
d–Av RF S	0.25	1.93775	0.13	0.8976
e–Av RF R	4.45	1.93775	2.30	0.0238

Parameter Estimates	*Percent Blue*			
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	64.967173	0.688031	94.42	<.0001
a–Arm RF S	1.0379707	0.788239	1.32	0.1909
b–Arm RF R	3.7456642	0.262746	14.26	<.0001
c–Av CS	-1.45518	0.788239	-1.85	0.0679
d–Av RF S	0.8308271	0.788239	1.05	0.2944
e–Av RF R	3.565094	0.788239	4.52	<.0001

Fig. 16 -- Main Factor Estimates

The remaining three factors, Armor Ranged Fire Range, Aviation Combat Strength, and Aviation Ranged Fire Range ('b', 'c', and 'e' respectively) were then analyzed for interaction effects. Since the Lack of Fit test passed for both MOE's, quadratic effects were not needed so axial trials were not accomplished. The entire 21 trial experiment with results is presented in Appendix C. Figure 17 shows that after all combinations were considered, only the two-way Ranged Fire Range interaction (b and e) appeared important. This interaction appeared somewhat significant for both MOE's. The two-factor interaction was used in both models.

```
┌──────────────────────────────────────────────────────────────────┐
│  ┌────────────────────┐        Blue                               │
│  │ Parameter Estimates │                                          │
│  └────────────────────┘                                           │
│   Term                  Estimate   Std Error   t Ratio  Prob>|t|   │
│   Intercept            160.22857   1.664784     96.25   <.0001     │
│   b-Arm RF R            9.8583333   0.63575     15.51    <.0001     │
│   e-Av RF R                 4.45    1.90725      2.3$    0.0216     │
│   b-Arm RF*e-Av RF      1.2416667   0.63575      1.95    0.0536     │
│                                                                    │
│                                                                    │
│  ┌────────────────────┐    Percent Blue                           │
│  │ Parameter Estimates │                                          │
│  └────────────────────┘                                           │
│   Term                    Estimate   Std Error  t Ratio  Prob>|t|  │
│   Intercept              64.967173   0.689359    94.24   <.0001    │
│   b-Arm RF R             3.7456642   0.263253    14.23   <.0001    │
│   c-Av CS                -1.45518    0.78976     -1.84   0.0684    │
│   e-Av RF R              3.565094    0.78976      4.51   <.0001    │
│   b-Arm RF*c-Av CS       0.0173752   0.263253     0.07   0.9475    │
│   b-Arm RF*e-Av RF       0.4953962   0.263253     1.88   0.0629    │
│   c-Av CS*e-Av RF        0.5774017   0.78976      0.73   0.4665    │
│   b-Arm RF*c-Av CS*e-Av RF  0.16074  0.263253     0.61   0.5429    │
└──────────────────────────────────────────────────────────────────┘
```

Fig. 17 -- Main Factor and Interaction Effects

Both MOE's are restricted by a maximum value. *Blue* is maximized when all of Red is eliminated. *Percent Blue* likewise is capped at 100%. A non-linear effect outside the

design area out to the maximum limit is very likely. Therefore, extrapolating information outside the design region is inadvisable. Table 5 shows the results of the regression. The sensitivity of the simulation's outcome is based on the figures below. A 10% increase of Armor's firing range will result in an improvement of Blue's score by about 10 points. Likewise, a nearly 4% increase of Blue's point share will result. This factor was the most significant in the research.

Table 5 – Analysis of Factors Regression Results

Factor	Blue	Percent Blue
Intercept	160	65
b - Armor RF Range	9.9	3.8
c - Aviation Combat Strength	N/A	-1.5
e - Aviation RF Range	4.5	3.5
Interaction be	1.2	0.5

The final coded regression models for each of the MOE's are listed in the following equations:

$$Blue = 160 + 9.9 \cdot b + 4.5 \cdot e + 1.2 \cdot b \cdot e \tag{32}$$

$$PercentBlue = 65 + 3.8 \cdot b - 1.5 \cdot c + 3.5 \cdot e + 0.5 \cdot b \cdot e \tag{33}$$

Both models perform very well against standard statistical tests. The models both have a normal residual distribution. All pertinent statistical information is included in Appendix D.

Each factor coefficient indicates the effect that the factor has on the simulation outcome. The simulation sensitivity is quantified with these coefficients. This particular simulation model is most sensitive to Armor Ranged Fire Range ('b'), followed by

Aviation Ranged Fire Range ('*e*'). Although Aviation Combat Strength ('*c*') is not as significant, a peculiar counter-intuitive behavior is indicated by its negative coefficient. As Blue Aviation Combat Strength is increased, Blue performance is reduced! This is discussed in the Validation section.

Screening results highlight those factors which the simulation model is most sensitive to. A follow-on analysis of those factors will provide further screening. Follow-on analysis also quantifies just how sensitive the simulation is to the remaining factors. With the results from Table 5, the simulation modeler should be most diligent with measurements concerning firing ranges of armor and aviation units. Further analysis of these factors should include those modifiers listed in Appendix A.

An uncoded model (a model which uses actual factor values) is nearly useless in this context. Each factor of this experiment was perturbed by a percentage value. These were coded into a consistent model where all factors were set ± 10% about their reference values. Therefore, all regression model coefficients are consistent with each other. An uncoded model would not be easy to use in this sensitivity context. For instance, consider a factor that is perturbed ± 10% around a reference value of 15, while another around 60. Individual regression coefficients would be consistent with each of these reference values. If these two factors had an equal effect on the simulation outcome, their coefficients would differ by a factor of four. As an example, the Equation (34) is the uncoded version of the *Blue* response in Equation (32).

$$Blue = 131 - 6.7 \cdot b - 4.7 \cdot e + 2.5 \cdot b \cdot e \qquad (34)$$

Unfortunately, this equation is only useful for predicting simulation outcomes. By inspection, there is nearly no way to get sensitivity information from this equation without first knowing the proper reference value for each factor! However, whichever method is used, coded or uncoded, a graphical plot of the resulting response surface will appear identical. These plots will provide a 'picture' that will allow a visual inspection of the regression model. These pictures will allow a sensitivity analysis simply by visual inspection. The following section will discuss the plots constructed for the regression models.

The Response Surface

A visual representation of the regression model can be useful in presenting results. Only two factors coupled with a response can be presented in a three dimensional plot. In the case of three factors, three separate plots presenting each pair combination are constructed. An example of a response surface plot is presented in Figure 18.
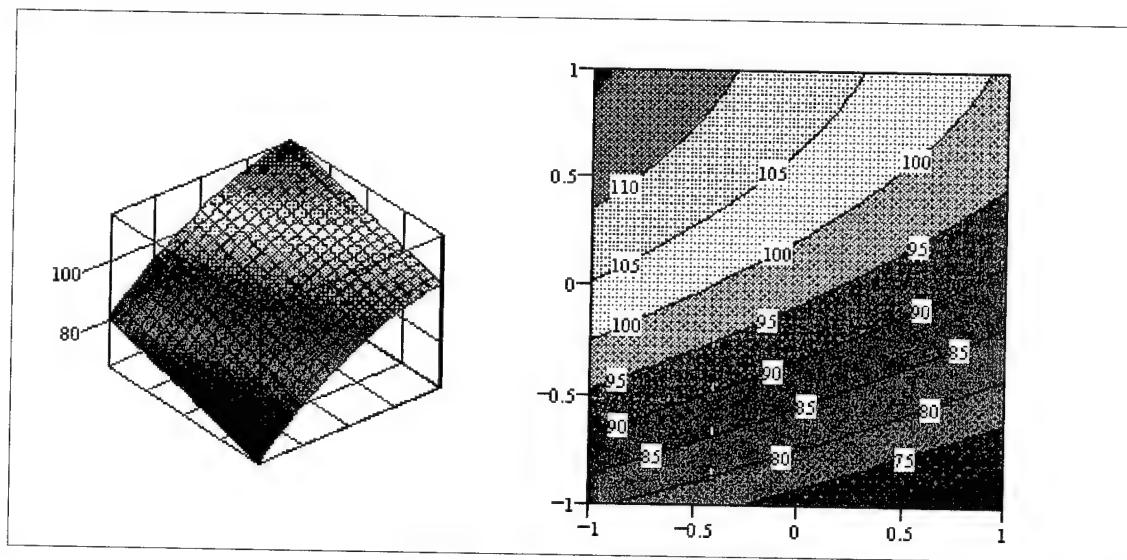
Fig 18 -- A Response Surface

Both the three-dimensional and two-dimensional plot of the same regression model are given in Figure 18. The two-dimensional plot may be more useful for sensitivity analysis. As the factor is increased up the Y axis, the response increases at a sizable rate which diminishes slightly. On the other hand, as the factor represented on the X axis is increased, the response slowly decreases at a constant rate. An inspection of the plot can give a relative 'feel' of the factor effects. Flat surface grades indicate those factors that the simulation outcome is insensitive or robust to. The simulation outcome does not depend on those factors. However, a factor that provides a hillside, particularly a steep one, indicates a factor the simulation outcome is especially sensitive to. Plots for each of the MOE's are presented in Appendix E.

Validation

Now that two models have been presented, a question of their credibility arises. Each of the models was constructed based on a set of input data. If the models are indeed accurate, they should predict any simulation outcome within the design region. Accurate predictions over the entire design region may mean the model could be a reasonable substitute for the simulation model itself. Two sources of independent data were used in the validation of each model.

The first set of validation data was generated by completing a full-factorial experiment with the three finalist factors. The other 20 factors were set at their reference values. Eight trials of five runs each produced 40 data points around the 'corners' of the

design region. The design was then supplemented with axial and center trials for a final total of 85 data points. This data provided a basic validation data set based only on those three factors which were used in each of the final regression models. The data used in the initial Plackett-Burman screening of factors were also used as another validation data set. However, this data set provided a tougher test for each of the final regression models because all 23 factors were varied, influencing each of the resulting validation data points. These biases should not have been individually important since the screening found them relatively insignificant. However, aggregate influence of these extra 20 'insignificant' factors exposed both models to a much more demanding test than the first set of validation data. If both of the final models are determined to be consistent with the screening data, then the factors used in those models are indeed the most significant. If the final models are moderately inconsistent, then the bias of the 20 unused factors most likely had some influence on the validation data. If the bias can be quantified, then the final models can be adjusted and applied accordingly.

Two techniques were used to compare the final models with the validation data sets. First, a regression surface was constructed using the same factors used by the final models. A comparison of the final models against the resulting regression surfaces will confirm if the factor coefficients are respectively consistent with one another. A second test used the predictions of the models against each individual validation data point. Mean Relative Error (MRE) showed if the models predict consistently high or low. MRE is presented as a positive or negative percentage. A good model will be consistent with the validation data and will have a very low MRE. MRE is calculated using Equation 35.

$$\text{MRE (Mean Relative Error)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right) \cdot 100\% \tag{35}$$

where
  n = Number of observations
  y_i = Observation from validation data.
  $\hat{y}_i$ = Observation prediction of the model.

A complimentary measure of the model's predictive ability is by use of Root Mean Square Error (RMSE). RMSE is an indication of how well a regression surface fits a set of data. The Coefficient of Determination (R^2) and RMSE are inversely related. A perfect fit will give an R^2 value of 1.0 and a RMSE of 0.0, and a bad fit an R^2 of a number near zero coupled with a very large RMSE. A regression surface based on a set of data will have some value for RMSE. The same surface compared against a different set of data will have another RMSE. Generally, a regression surface will have a lower RMSE against the data from which it was constructed verses an independent set of data. Regardless, if the two RMSE values are similar, then a conclusion can be drawn that the regression surface closely fits the independent data. RMSE is calculated with the following formula:

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}} \tag{36}$$

With a combination of low MRE and consistent RMSE, a regression model can be accepted as a accurate predictor. The actual point where the model is considered

84

adequate is a subjective decision. The user's tolerance to model error should define the satisfactory validation criteria.

Validation Results. After collecting and organizing the validation data, a least square regression surface was constructed for each. All validation data information is available in Appendix F. Table 6 summarizes the coefficients of the original surfaces and the validation surfaces. Notice that several factors proved to be less than 95% significant in the validation models. That does not mean the final model's terms in turn lack significance, but simply that the respective validation data set was not sensitive to those factors. Low values for R^2 (Appendix F) are not due to a bad fit, but to the high variance from numerous trial runs. Using average trial values will give similar results and a higher R^2 (over .95), but with much fewer degrees of freedom for other statistical calculations. Overall, although each model is based on a different data set, the three response surfaces are very consistent.

Table 6 -- Response Surface Coefficient Comparison

	Final Model (Original Data)	Validation Model (Full-Factorial Data)	Validation Model (Screening Data)
MOE: *Blue*			
Intercept	160	160	151
***b* - Arm RF R**	9.9	10.0	8.9
***e* - Avn RF R**	4.5	Insignificant**	4.9*
Interaction *be*	1.2	1.0*	Insignificant**
MOE: *Percent Blue*			
Intercept	65	65	60
***b* - Arm RF R**	3.7	3.5	3.3
***c* - Avn CS**	-1.5	-1.1*	-3.7
***e* - Avn RF R**	3.6	1.7	3.3
Interaction *be*	0.5	0.5	0.5*

* Did not meet 95% Significance Criteria
** Completely Insignificant -- Did not meet 50% Significance Criteria

The outcome of *Blue's* score is consistently very sensitive to the Armor's Ranged Fire

Range capability. Aviation's Ranged Fire Range capability may indeed be somewhat

important, although that was not confirmed by either validation data set. However,

relative performance of the Blue aggressors as shown by the *Percent Blue* MOE is rather

sensitive to all three factors. Notice that the counter-intuitive behavior of Aviation's

Combat Strength is confirmed by both validation sets. Both Ranged Fire Range factors

proved consistently significant for *Percent Blue* across all three data sets.

The predictive ability of each of the two models was then measured against the

validation data. Table 7 summarizes the findings:

Table 7 – Predictive Comparisons

	Final Model (Original Data)	Validation Model (Full-Factorial Data)	Validation Model (Screening Data)
MOE: *Blue*			
MRE	N/A (0%)	+1.6%	+9.7%
RMSE	17.1	18.9	19.5
MOE: *Percent Blue*			
MRE	N/A (0%)	-0.1%	+10.5%
RMSE	7.0	6.2	6.3

The final models do very well against the first set of validation data. RME of 1.6%

and -0.1% indicate the validation data from the full-factorial data set almost perfectly

brackets both regression models. Both RMSE's indicate the trend of the data nearly

matches the trend of the original data. Both of the final models fare very well against the

full-factorial data set, which really was not unexpected. The final models were

constructed from a Resolution V experiment with five factors influencing the original data set. The full-factorial data set was comprised of data where only the three final factors were manipulated. No bias from the 20 'insignificant' factors was introduced.

Both of the final models tended to predict about 10% high against the more biased screening data. Apparently, the screening data may not only be biased by the extra factors, but also by non-orthogonal factor settings. The Plackett-Burman design does not randomly change factor values from trial to trial. The design is orthogonal only if used in its entirety. The validation screening data was separated into eight non-orthogonal subsets. That introduces a bias which may explain why the screening data shows a lack of performance for the Blue aggressors. Although predictions tended to be high in both cases, the RMSE produced by both validation data sets compare favorably with the original data's regression model.

The RMSE was better with both validation data sets when *Percent Blue* was used as an MOE. In both cases, the validation data produced an unusually improved consistency when placed against the final *Percent Blue* model. Most likely the original data had a higher variance with respect to the *Percent Blue* MOE. The same trend did not appear with the final *Blue* MOE model.

Overall, both of the final models predict very well against data sets involving the three final factors. A data set where all 23 factors are randomly altered may offer a better validation result than the original Plackett-Burman screening experiment data set. Until that is resolved, both of the final models may be suspect in predicting outcomes of simulation runs where all 23 factors are considered. However, both models prove themselves as accurate 'mappings' of the simulation's behavior and can be used as

reliable sensitivity analysis tools. The validated *Percent Blue* MOE model indicated the simulation responds to Aviation Combat Strength in a counter-intuitive manner. The next section investigates this behavior.

Simulation in Question -- The Counter-Intuitive Result. The last issue confronted was the counter-intuitive result involving Blue Army's Aviation Combat Strength ('*c*'). As strength increased, Blue's performance went down significantly. Red performance also increased accordingly. Separate simulation trials were run where the factor was altered in increments of 10% from -30% to +30% of the reference value.

Over the large span, as Aviation Combat Strength increased, Blue's overall battlefield performance decreased. Blue outscored Red 71% to 29% (of total score) when strength was set at -30%. At +30%, Blue outscored Red 63% to 37%. In general, Blue loses effectiveness as Aviation Combat Strength increases. This wide span confirms the odd behavior detected by all of the regression metamodels. Figure 19 shows a plot off all simulation runs in terms of the *Percent Blue* MOE.
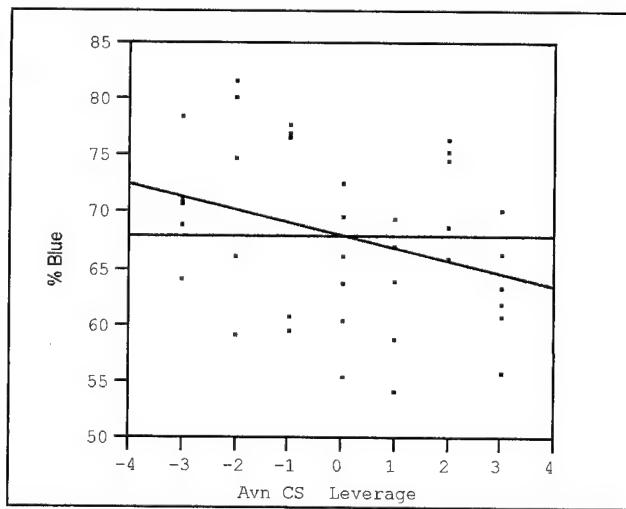
Fig. 19 – Aviation Combat Strength Trend

Equation (37) shows the overall effect of the factor. This equation is based on a separate set of simulation trials in which only Blue's Aviation Combat Strength was altered. The result is remarkably close to the respective values in the final *Percent Blue* regression model.

$$PercentBlue = 68 - 1.1 \cdot c \qquad (37)$$

By visually observing the simulation runs in progress, a trend explaining the behavior appeared. As expected, the aggressiveness of Blue Aviation intensified as Combat Strength increased. However, Blue Aviation units were putting themselves into much greater risk while working around the front lines. More intrusions across the line resulted in many more Red shots hitting their mark. As strength and in turn aggressiveness increased, Blue Aviation units were eliminated despite their greater strength. Helicopters stayed back when strength was set low, and blindly attacked when strength was set high. For example, of the four aviation units, an average of 1.5 unit were lost when strength was 30% lower than the reference value. On the other hand, an average of 3.3 units were lost when strength was set 30% higher! Overall, the *Blue* MOE remained somewhat constant, but Red's score increased proportionally to the helicopter units killed. Therefore, the *Percent Blue* MOE showed the effect most clearly.

Other unit types also had a similar trend. Refer to the screening experiment for the *Blue* MOE (Appendix B). Most other units types had a slightly negative trend as their combat strength was increased. Blue Aviation units were by far the highest value target for Red units (15 points per kill), so the negative Blue Aviation trend was highlighted with the MOE's used.

The simulation model has indicated an interesting phenomena. The artificial intelligence of the simulation model elects to increase aggressiveness as combat strength is increased. Some say this may be a tendency of most human leaders. The simulation model points out that maybe aggressiveness should be left alone, allowing weapon systems a longer life span. In turn, the weapon system can accumulate more kills during the long run. In conclusion, the simulation model may be accurately reflecting a combat tendency, and should not be disregarded. Aggressiveness should be carefully amplified as strength is increased.

In summary, both of the final models are very good at sensitivity analysis. Using only three of the original 23 factors, the regression models also provide a useful mapping of the simulation's behavior. Counter-intuitive simulation behavior was highlighted and investigated. The predictive ability of both models is shown to be reasonable. However, the deterministic nature of a regression model means it may not be beneficial as a short-sighted predictor of a highly stochastic simulation. On the other hand, a prediction from a regression model should be very useful for long range 'expected value' information. Isolating counter-intuitive behavior proved to be a very valuable benefit of the regression model. New simulation models may benefit from a regression analysis for such a reason. Both of the final models proved themselves as valuable mappings of the combat simulation.

VI. Conclusion / Recommendations

Objective Review

The primary objective of this thesis was to measure the sensitivity of a combat simulation's outcome to various weapon system capabilities. Since weapon capabilities of established combat simulations are hard-coded, a more accessible combat simulation was used. *Empire II: The Art of War*™, a computer wargame was used as a substitute. The game offered a computer verses computer option, thus turning the wargame into a 'combat simulation'. After designing a battle scenario, weapon system capabilities on a single side were manipulated. Those manipulated factors indeed made a difference to the simulation's battle outcome. Using Response Surface Methodology, those factors that had the most influence were isolated. As shown in the Literature Review, RSM was a sound technique for sensitivity analysis.

Research Summary

Previous theses efforts took an established combat simulation, a given scenario, and a given set of weapon systems, then manipulated a side's force structure. Their objective was to isolate those weapon systems which had the most influence to the outcome. However, in this research the battlefield force structure remained constant. One side's weapon system capabilities were manipulated to identify how much significance they had on the outcome. Five separate weapon systems were simulated, each having several unique capabilities. A total of 23 factors were identified as potentially important to the

91

simulation outcome. A sensitivity analysis was then performed using RSM techniques. Only those factors which the simulation outcome was most sensitive remained for use in the final regression metamodel.

In the final RSM metamodel, only three of the initial 23 factors were required to adequately describe the simulation's behavior. The results showed that a simple regression model can indeed be constructed based on a limited number of simulation trials. In the end, the outcome's sensitivity was defined by each factor's coefficient. The metamodel may also be used as a quick-turn deterministic 'simulation'. This kind of application could be used for a general 'first-look' analysis. A simple metamodel may conveniently and adequately serve as a substitute for actual simulation trials.

Significance of Results

In practice, a similar procedure could be used with an actual combat simulation. For simulation accuracy, those highlighted factors should be measured with the most care as they are updated in the simulation model. If these factors are indeed important to the simulation outcome, then these factors are likely to be just as important on the actual battlefield. Commanders should be aware of what they are. On the other hand, these commanders should also be made aware of those factors that do not make a difference. Actual resources spent on those factors may be excessive.

The acquisition process may be very interested in finding those factors that are most significant. Many times, a disproportionate amount of money is spent upgrading those factors that would make little difference in a battle. Highlighting those most significant

factors of a combat simulation should prove as very useful information for decisions concerning the allocation of resources destined for upgrades.

Military professionals have recently used commercial wargames for battle insights (Dunnigan, 1992: 234). Even though this particular wargame is by no means a substitute for a complex combat simulation, it can still actually show which weapon capabilities may be most important. A similar effort with an actual combat simulation would confirm if indeed a wargame like this one is consistent with the multi-million dollar simulation model. In his Wargames Handbook, noted wargame enthusiast James F. Dunnigan states:

> Until recently, professional wargaming tended to go after the less than 1 percent deviation in accuracy and consume years and millions of dollars in the process. Indeed, the process usually overcame the search for a solution, often leading to a lost, and failed, project. Commercial manual wargames can generally achieve a '5 percent solution'. Put a wargame on a PC, and you get a nearly 1 percent solution. But it took the military a while to catch on to this cheaper solution sitting on the shelves of a local game store. (Dunnigan, 1992: 236)

If wargame behavior is indeed consistent with established combat simulations, Mr. Dunnigan would have his point about the value of commercial wargames validated.

Recommendation for Further Research. An established combat simulation and a commercially produced computer wargame could be compared for consistency of results. For a given scenario, if the commercial wargame is found to be consistent with the established combat simulation, an argument where 'cheap' wargames could be used as an alternative could be defended. However, a combat simulation which could be altered must be found. Many accredited simulation models may not be available for this kind of research.

Possible Improvements

The methodology used in this research has been used several times before as mentioned in the Literature review. The primary flaw with the method is in the possible over simplification of a process. This particular metamodel will serve as a tool for only this particular scenario. The metamodel should not be used to predict results or trends in unrelated scenarios. Multiple experiments expanding into several scenarios should be considered for a conglomeration of metamodels that together could be used for wide ranging decisions. However, such a tasking will be time consuming since a multitude of simulation runs would be required.

Recommendation for Further Research. If these metamodels could be constructed with a consistent RSM technique, all could be combined into one generic model. As long as each factor is considered equally in all of the models, the coefficients may be averaged into a mean parameter. Since each coefficient is actually a normal random variable, a mean normal distribution may be found each parameter. A stochastic spreadsheet metamodel can produce results with all scenarios considered. In other words, an aggregated quick-turn metamodel may be constructed piece by piece over time. As new scenarios are added and others dropped, the overall metamodel can be updated accordingly. Certain factors will be identified as significant for the overall picture, while others will be highlighted for individual scenarios. Not unlike an object-oriented simulation, each individual metamodel can be 'plugged in' as needed into the aggregated model.

Summary

Applying the RSM technique to map out a simulation model is extremely useful. This research utilized the technique for sensitivity analysis purposes. The metamodel mapping is beneficial for other applications such as quick-turn analysis where simply 'plugging in' factor values will produce a prediction of the simulation result, not to mention an actual result. Because of tremendous variability, combat modeling may be one of the more risky areas to use this somewhat deterministic method. As long as the limits of the metamodel are recognized, the risk should be acceptable. In processes where results are consistent in nature such as a manufacturing process, a regression metamodel can be very helpful, not only in spotting trends but also as an accurate predictor of results.

A major benefit of the RSM model was its ability to highlight counter-intuitive simulation behavior. These behaviors can be further investigated and repaired if required. Although not an expected result, the RSM model proved to serve as a verification tool for the simulation model itself.

Overall, Response Surface Methodology is a proven technique which can be used to map a process to the point where decision makers can see the 'big picture' without reckless trial and error. By using RSM, this research objective was easily met.

Appendix A: Entity Reference Settings

Empire II Technology Report
Reference Scenario

1. Rules

Turn Length	72 min
Combat Damage Effectiveness	50%
Combat Variance	0%
Entrenchment Modifier	5
Transit Modifier	-5
Supply Rules	No
Morale Rules	Yes
Sighting Rules	Yes
Detection Range	30
Identification Range	10

2. Forces

Combat Efficiency

Blue	_0_
Red	0

3. Unit Type Summary

Blue	Speed	RF Cost	Max Strength	Kill Pt	Morale Pt	RF Strength	RF Accuracy
Artillery	_10_	3	25	0	10	_25_	_25%_
Mech Inf	_15_	6	40	0	10	_15_	_60%_
Armor	_15_	6	70	0	10	_25_	_70%_
Aviation	_60_	40	65	0	30	_35_	_70%_
CAS	_30_	30	100	0	0	_30_	_50%_

Red							
Light Inf	2	2	50	0	0	8	80%
C2	2	2	90	0	0	30	80%
Artillery	10	3	25	0	10	25	25%
Mech Inf	15	6	40	0	10	15	60%
Armor	15	6	70	0	10	25	70%
Aviation	60	40	65	0	30	35	70%

Blue Artillery

(** Selected Variable)

Speed **	*10*			Kill Break Point	0
Transit Speed	20			Morale Break Point	10
RF Cost	3			Max Combat Strength **	*25*
RF Type	Artillery			Max RF Strength **	*25*
RF Range **	*15*			RF Depletion	0
				RF Accuracy **	*25%*
				Able to Entrench?	Yes
				Zone of Control?	Yes

Sighting

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	15	Red Artillery	10	Red Mech Inf	7
Red Armor	10	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	10
Village	10	Clear	6	Forest	10
Broken Ground	12	Mountain	12	Desert	8
Arctic	8	Swamp	12	Beach	10
River	4	Stream	2	Road	5

Movement Costs (Transit)

Deep Sea	No	Shallow Sea	No	City	8
Village	4	Clear	2	Forest	6
Broken Ground	8	Mountain	10	Desert	6
Arctic	6	Swamp	10	Beach	6
River	4	Stream	2	Road	4

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-3
Broken Ground	0	Mountain	-3	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	0
Broken Ground	12	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	1	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	-5	Red Artillery	3	Red Mech Inf	2
Red Armor	2	Red Aviation	-7		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	No
Red C2	3	Red Artillery	0	Red Mech Inf	-3
Red Armor	-4	Red Aviation	-7		

Blue Mechanized Infantry
(** Selected Variable)

Speed **	15	Kill Break Point	0
Transit Speed	NA	Morale Break Point	10
RF Cost	6	Max Combat Strength **	40
RF Type	Guns	Max RF Strength **	15
RF Range **	4	RF Depletion	0
		RF Accuracy **	60%
		Able to Entrench?	Yes
		Zone of Control?	Yes

Sighting

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	15	Red Artillery	5	Red Mech Inf	5
Red Armor	5	Red Aviation	5		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	8
Village	6	Clear	2	Forest	6
Broken Ground	6	Mountain	8	Desert	8
Arctic	8	Swamp	8	Beach	6
River	4	Stream	2	Road	6

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-2
Broken Ground	0	Mountain	-3	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	0
Broken Ground	3	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	2	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	-7	Red Artillery	1	Red Mech Inf	2
Red Armor	2	Red Aviation	1		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	No
Red C2	1	Red Artillery	0	Red Mech Inf	0
Red Armor	-1	Red Aviation	2		

Blue Armor

(** Selected Variable)

Speed **	15	Kill Break Point		0
Transit Speed	NA	Morale Break Point		10
RF Cost	6	Max Combat Strength **		70
RF Type	Incendiary	Max RF Strength **		25
RF Range **	3	RF Depletion		0
		RF Accuracy **		70%
		Able to Entrench?		Yes
		Zone of Control?		Yes

Sighting

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	15	Red Artillery	10	Red Mech Inf	5
Red Armor	7	Red Aviation	3		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	8
Village	6	Clear	2	Forest	6
Broken Ground	6	Mountain	8	Desert	8
Arctic	8	Swamp	8	Beach	6
River	4	Stream	2	Road	7

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-3
Broken Ground	0	Mountain	-4	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	0
Broken Ground	3	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	1	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	-6	Red Artillery	3	Red Mech Inf	3
Red Armor	3	Red Aviation	-4		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	No
Red C2	2	Red Artillery	1	Red Mech Inf	1
Red Armor	1	Red Aviation	-3		

Blue Aviation
(** Selected Variable)

Speed **	*60*		Kill Break Point	0
Transit Speed	NA		Morale Break Point	30
RF Cost	40		Max Combat Strength **	*65*
RF Type	Rockets		Max RF Strength **	*35*
RF Range **	*15*		RF Depletion	0
			RF Accuracy **	*70%*
			Able to Entrench?	No
			Zone of Control?	No

Sighting

Red Lgt Inf	-5	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	15	Red Artillery	8	Red Mech Inf	5
Red Armor	5	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	2
Village	2	Clear	2	Forest	4
Broken Ground	2	Mountain	6	Desert	2
Arctic	2	Swamp	2	Beach	2
River	NA	Stream	NA	Road	NA

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-1
Broken Ground	0	Mountain	-2	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	-3
Broken Ground	3	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	1	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	-5	Red Artillery	2	Red Mech Inf	3
Red Armor	2	Red Aviation	-7		

Target Ranged Fire Adjustments

Red Lgt Inf	-1	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	No
Red C2	-7	Red Artillery	0	Red Mech Inf	-1
Red Armor	0	Red Aviation	-7		

Blue CAS
(** Selected Variable)

Speed	30	Kill Break Point	0
Transit Speed	NA	Morale Break Point	0
RF Cost	30	Max Combat Strength	100
RF Type	Jet Bomber	Max RF Strength **	*30*
RF Range	130	RF Depletion	0
		RF Accuracy **	*50%*
		Able to Entrench?	Yes
		Zone of Control?	No

Sighting

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	76	Red Artillery	70	Red Mech Inf	70
Red Armor	70	Red Aviation	5		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	No
Village	No	Clear	Placed	Forest	No
Broken Ground	No	Mountain	No	Desert	No
Arctic	No	Swamp	No	Beach	No
River	NA	Stream	NA	Road	NA

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-3
Village	-5	Clear	0	Forest	-6
Broken Ground	0	Mountain	-6	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	-3
Broken Ground	3	Mountain	-2	Desert	3
Arctic	3	Swamp	-2	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	1	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	0
Red C2	-8	Red Artillery	2	Red Mech Inf	1
Red Armor	2	Red Aviation	-7		

Target Ranged Fire Adjustments

Red Lgt Inf	-1	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	No
Red C2	-7	Red Artillery	0	Red Mech Inf	-1
Red Armor	0	Red Aviation	-7		

Red Light Infantry

Speed	2		Kill Break Point	0
Transit Speed	NA		Morale Break Point	0
RF Cost	2		Max Combat Strength	50
RF Type	Guns		Max RF Strength	8
RF Range	2		RF Depletion	0
			RF Accuracy	80%
			Able to Entrench?	Yes
			Zone of Control?	No

Sighting

Red Lgt Inf	0	Blue Artillery	5	Blue Mech Inf	5
Blue Armor	5	Blue Aviation	5	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	2
Village	No	Clear	No	Forest	No
Broken Ground	No	Mountain	No	Desert	No
Arctic	No	Swamp	No	Beach	No
River	NA	Stream	NA	Road	NA

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	1
Village	0	Clear	0	Forest	0
Broken Ground	0	Mountain	0	Desert	0
Arctic	0	Swamp	0	Beach	0
River	1	Stream	0		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	0
Village	0	Clear	0	Forest	0
Broken Ground	0	Mountain	0	Desert	0
Arctic	0	Swamp	0	Beach	0
River	0	Stream	0		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	0	Blue Artillery	2	Blue Mech Inf	4
Blue Armor	4	Blue Aviation	1	Blue CAS	0
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	1	Blue Mech Inf	1
Blue Armor	1	Blue Aviation	3	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Red Command and Control

Speed	2		Kill Break Point	0
Transit Speed	NA		Morale Break Point	0
RF Cost	2		Max Combat Strength	90
RF Type	Guns		Max RF Strength	30
RF Range	2		RF Depletion	0
			RF Accuracy	80%
			Able to Entrench?	Yes
			Zone of Control?	Yes

Sighting

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	0	Blue Aviation	0	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	Place	Shallow Sea	Place	City	Place
Village	Place	Clear	Place	Forest	Place
Broken Ground	Place	Mountain	Place	Desert	Place
Arctic	Place	Swamp	Place	Beach	Place
River	NA	Stream	NA	Road	NA

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	0
Village	0	Clear	0	Forest	0
Broken Ground	0	Mountain	0	Desert	0
Arctic	0	Swamp	0	Beach	0
River	1	Stream	0		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	2
Village	0	Clear	0	Forest	0
Broken Ground	0	Mountain	0	Desert	0
Arctic	0	Swamp	0	Beach	0
River	0	Stream	0		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	0	Blue Artillery	2	Blue Mech Inf	2
Blue Armor	2	Blue Aviation	-2	Blue CAS	0
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Target Ranged Fire Adjustments

Red Lgt Inf	2	Blue Artillery	2	Blue Mech Inf	2
Blue Armor	2	Blue Aviation	7	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Red Artillery

Speed	10	Kill Break Point	0
Transit Speed	20	Morale Break Point	10
RF Cost	3	Max Combat Strength	25
RF Type	Artillery	Max RF Strength	25
RF Range	15	RF Depletion	0
		RF Accuracy	25%
		Able to Entrench?	Yes
		Zone of Control?	Yes

Sighting

Red Lgt Inf	0	Blue Artillery	10	Blue Mech Inf	7
Blue Armor	10	Blue Aviation	0	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	10
Village	10	Clear	6	Forest	10
Broken Ground	12	Mountain	12	Desert	8
Arctic	8	Swamp	12	Beach	10
River	4	Stream	2	Road	5

Movement Costs (Transit)

Deep Sea	No	Shallow Sea	No	City	8
Village	4	Clear	2	Forest	6
Broken Ground	8	Mountain	10	Desert	6
Arctic	6	Swamp	10	Beach	6
River	4	Stream	2	Road	4

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-3
Broken Ground	0	Mountain	-3	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	0
Broken Ground	3	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	1	Blue Artillery	3	Blue Mech Inf	2
Blue Armor	2	Blue Aviation	-7	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	-3
Blue Armor	-4	Blue Aviation	-7	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Red Mechanized Infantry

Speed	15		Kill Break Point	0
Transit Speed	NA		Morale Break Point	10
RF Cost	6		Max Combat Strength	40
RF Type	Guns		Max RF Strength	15
RF Range	4		RF Depletion	0
			RF Accuracy	60%
			Able to Entrench?	Yes
			Zone of Control?	Yes

Sighting

Red Lgt Inf	0	Blue Artillery	5	Blue Mech Inf	5
Blue Armor	5	Blue Aviation	5	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	8
Village	6	Clear	2	Forest	6
Broken Ground	6	Mountain	8	Desert	8
Arctic	8	Swamp	8	Beach	6
River	4	Stream	2	Road	6

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-2
Broken Ground	0	Mountain	-3	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	0
Broken Ground	3	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	2	Blue Artillery	1	Blue Mech Inf	2
Blue Armor	2	Blue Aviation	1	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	0
Blue Armor	-1	Blue Aviation	2	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Red Armor

Speed	15		Kill Break Point	0
Transit Speed	NA		Morale Break Point	10
RF Cost	6		Max Combat Strength	70
RF Type	Incendiary		Max RF Strength	25
RF Range	3		RF Depletion	0
			RF Accuracy	70%
			Able to Entrench?	Yes
			Zone of Control?	Yes

Sighting

Red Lgt Inf	0	Blue Artillery	10	Blue Mech Inf	5
Blue Armor	7	Blue Aviation	3	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	8
Village	6	Clear	2	Forest	6
Broken Ground	6	Mountain	8	Desert	8
Arctic	8	Swamp	8	Beach	6
River	4	Stream	2	Road	7

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-3
Broken Ground	0	Mountain	-4	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	0
Broken Ground	3	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	1	Blue Artillery	3	Blue Mech Inf	3
Blue Armor	3	Blue Aviation	-4	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	1	Blue Mech Inf	1
Blue Armor	1	Blue Aviation	-3	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Red Aviation

Speed	60			Kill Break Point	0
Transit Speed	NA			Morale Break Point	30
RF Cost	40			Max Combat Strength	65
RF Type	Rockets			Max RF Strength	35
RF Range	15			RF Depletion	0
				RF Accuracy	70%
				Able to Entrench?	Yes
				Zone of Control?	No

Sighting

Red Lgt Inf	0	Blue Artillery	8	Blue Mech Inf	5
Blue Armor	5	Blue Aviation	0	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Movement Costs (Speed)

Deep Sea	No	Shallow Sea	No	City	2
Village	2	Clear	2	Forest	4
Broken Ground	2	Mountain	6	Desert	2
Arctic	2	Swamp	2	Beach	2
River	NA	Stream	NA	Road	NA

Terrain Combat Adjustments

Deep Sea	0	Shallow Sea	0	City	-2
Village	-3	Clear	0	Forest	-1
Broken Ground	0	Mountain	-2	Desert	0
Arctic	0	Swamp	-2	Beach	0
River	2	Stream	1		

Terrain Ranged Fire Adjustments

Deep Sea	0	Shallow Sea	0	City	-5
Village	-3	Clear	1	Forest	-3
Broken Ground	3	Mountain	2	Desert	3
Arctic	3	Swamp	1	Beach	3
River	2	Stream	1		

Attacker vs Defender Combat Adjustments

Red Lgt Inf	1	Blue Artillery	2	Blue Mech Inf	1
Blue Armor	2	Blue Aviation	-7	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Target Ranged Fire Adjustments

Red Lgt Inf	0	Blue Artillery	0	Blue Mech Inf	-1
Blue Armor	0	Blue Aviation	-7	Blue CAS	No
Red C2	0	Red Artillery	0	Red Mech Inf	0
Red Armor	0	Red Aviation	0		

Appendix B: Screening Experiment

Plackett-Burman Design (N = 23)

Trial	Act Move	Act CS	Act RF S	Act RF R	Act RF A	Mech Move	Mech CS	Mech RF S	Mech RF R	Mech RF A	Arm Move	Arm CS	Arm RF S	Arm RF R	Arm RF A	Av Move	Av CS	Av RF S	Av RF R	Av RF A	CAS RF S	CAS RF A	Efficiency
1	1	1	1	1	1	1	1	1	2.5	1	1	1	1	3	1	1	1	1	1	1	1	1	1
2	-1	1	1	1	1	-1	1	-1	2.5	1	-1	-1	1	3	-1	-1	1	-1	1	1	1	1	-1
3	-1	-1	1	1	1	1	-1	1	-2.5	1	1	-1	-1	3	1	1	-1	1	1	-1	1	-1	-1
4	-1	-1	-1	1	1	1	1	-1	2.5	-1	1	1	-1	-3	-1	1	-1	-1	1	1	1	1	-1
5	-1	-1	-1	-1	1	1	1	1	-2.5	1	-1	1	1	-3	1	-1	1	-1	-1	1	1	1	-1
6	-1	-1	-1	-1	-1	1	1	1	2.5	-1	1	-1	1	3	-1	1	-1	1	-1	-1	1	1	-1
7	1	-1	-1	-1	-1	-1	1	1	2.5	1	-1	1	-1	3	1	-1	1	-1	1	-1	1	1	-1
8	-1	1	-1	-1	-1	-1	-1	1	2.5	-1	1	-1	1	-3	1	1	1	1	-1	1	1	1	-1
9	1	-1	1	-1	-1	-1	-1	-1	2.5	1	1	1	-1	3	-1	1	-1	1	1	1	1	1	-1
10	-1	-1	-1	1	-1	-1	-1	-1	-2.5	1	-1	1	1	-3	1	-1	1	1	1	1	1	1	-1
11	-1	-1	1	-1	1	-1	-1	-1	-2.5	-1	1	-1	1	3	-1	-1	-1	1	1	1	1	1	1
12	1	-1	-1	1	-1	1	-1	-1	-2.5	1	1	-1	-1	3	1	1	1	1	1	1	1	1	1
13	-1	-1	-1	-1	1	-1	1	-1	-2.5	-1	1	1	-1	-3	1	-1	-1	1	1	1	1	1	1
14	-1	1	-1	-1	-1	1	-1	1	-2.5	-1	-1	1	1	-3	-1	1	1	1	1	1	1	1	1
15	1	-1	1	-1	-1	-1	1	-1	2.5	-1	-1	-1	1	-3	1	-1	1	1	1	1	1	1	1
16	1	1	-1	1	-1	-1	-1	1	-2.5	1	-1	-1	-1	-3	1	-1	1	1	1	1	1	1	1
17	1	1	1	-1	1	1	-1	-1	2.5	1	1	-1	-1	3	-1	1	1	1	1	1	1	1	1
18	-1	1	1	1	-1	-1	1	-1	-2.5	-1	-1	1	-1	-3	1	-1	1	1	1	1	1	1	1
19	1	-1	1	1	1	1	-1	1	-2.5	1	-1	-1	1	-3	-1	1	-1	1	1	1	1	1	1
20	-1	1	-1	1	1	-1	1	-1	2.5	-1	1	-1	-1	3	1	1	1	1	1	1	1	1	1
21	-1	-1	1	-1	1	1	-1	-1	-2.5	1	1	1	-1	-3	-1	-1	1	1	1	1	1	1	1
22	1	-1	-1	1	-1	1	1	-1	2.5	-1	-1	1	1	3	1	1	-1	1	1	1	1	1	1
23	1	1	-1	-1	1	-1	1	1	-2.5	-1	-1	1	1	-3	-1	-1	1	1	1	1	1	1	1
24	1	1	1	1	-1	-1	-1	1	2.5	-1	-1	1	1	-3	-1	1	1	1	1	1	1	1	-1

Plackett-Burman Screen—Run Summaries

Trial	Run	Blue	Red	Percent Blue	Trial	Run	Blue	Red	Percent Blue
1	1	219	22	90.9	13	37	183	89	67.3
	2	212	51	80.6		38	177	99	64.1
	3	191	76	71.5		39	178	72	71.2
2	4	129	83	60.8	14	40	140	79	63.9
	5	158	69	69.6		41	166	108	60.6
	6	190	55	77.6		42	162	100	61.8
3	7	213	66	76.3	15	43	91	172	34.6
	8	179	89	66.8		44	85	138	38.1
	9	155	78	66.5		45	108	167	39.3
4	10	162	103	61.1	16	46	85	114	42.7
	11	134	160	45.6		47	110	115	48.9
	12	142	151	48.5		48	104	121	46.2
5	13	85	121	41.3	17	49	119	103	53.6
	14	62	149	29.4		50	92	161	36.4
	15	114	113	50.2		51	126	111	53.2
6	16	206	68	75.2	18	52	123	122	50.2
	17	167	73	69.6		53	113	98	53.6
	18	209	64	76.6		54	111	100	52.6
7	19	157	61	72.0	19	55	112	110	50.5
	20	192	68	73.8		56	116	122	48.7
	21	172	38	81.9		57	106	173	38.0
8	22	149	87	63.1	20	58	162	80	66.9
	23	176	87	66.9		59	181	75	70.7
	24	188	90	67.6		60	133	113	54.1
9	25	159	114	58.2	21	61	131	114	53.5
	26	120	130	48.0		62	185	110	62.7
	27	147	102	59.0		63	155	128	54.8
10	28	97	113	46.2	22	64	170	135	55.7
	29	101	150	40.2		65	144	117	55.2
	30	120	116	50.8		66	170	87	66.1
11	31	204	28	87.9	23	67	119	175	40.5
	32	209	53	79.8		68	98	143	40.7
	33	194	95	67.1		69	104	155	40.2
12	34	179	53	77.2	24	70	146	120	54.9
	35	178	103	63.3		71	138	169	45.0
	36	165	130	55.9		72	139	99	58.4

Normal Probability Plots
Plackett-Burman Screen

Blue
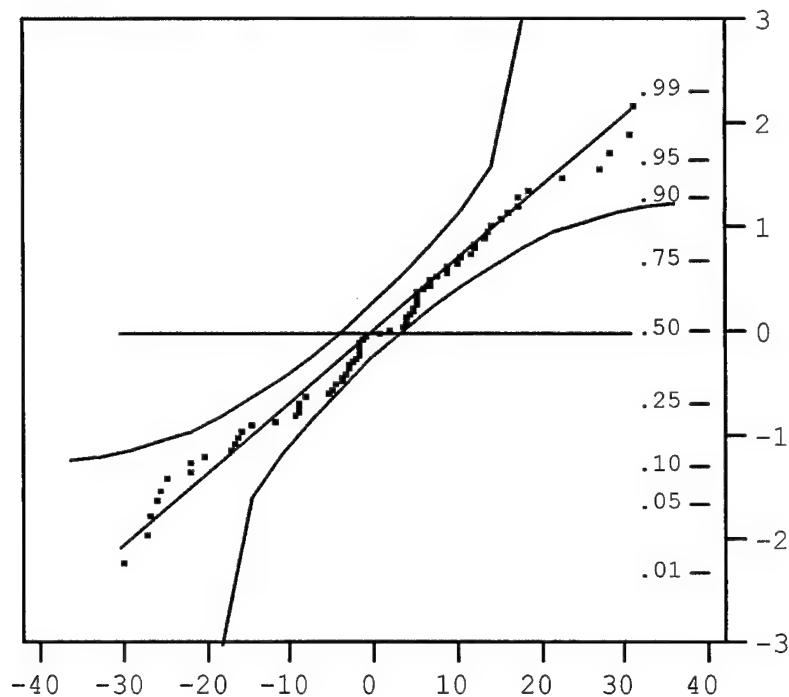
Percent Blue
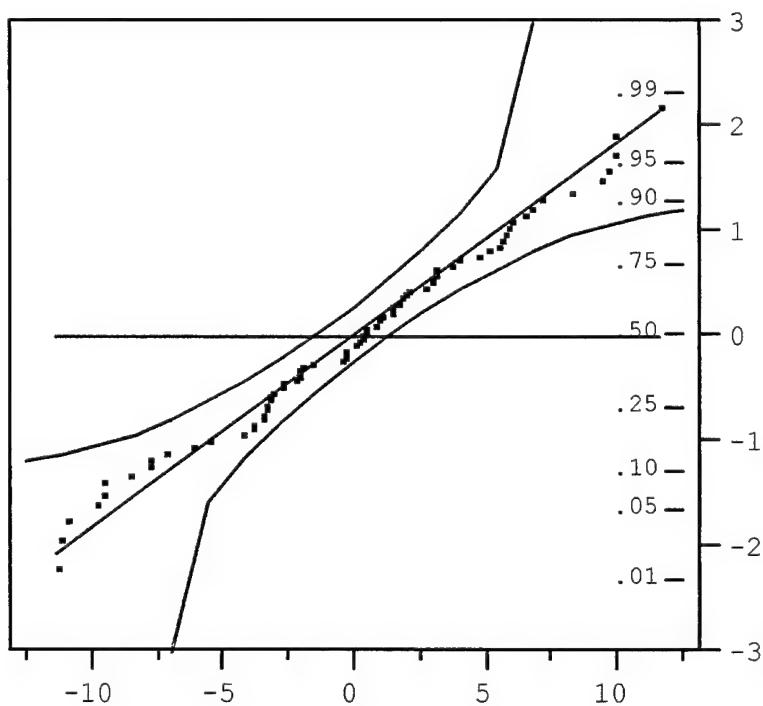
Blue Screen

Summary of Fit

RSquare	0.855875
RSquare Adj	0.786815
Root Mean Square Error	17.54874
Mean of Response	147.4444
Observations (or Sum Wgts)	72

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	147.44444	2.068139	71.29	<.0001
Art M	-0.277778	2.068139	-0.13	0.8937
Art CS	0.4444444	2.068139	0.21	0.8308
Art RF S	0.5277778	2.068139	0.26	0.7997
Art RF R	-2.333333	2.068139	-1.13	0.2648
Art RF A	2.3888889	2.068139	1.16	0.2538
Mech M	5.3888889	2.068139	2.61	0.0122
Mech CS	-2.75	2.068139	-1.33	0.1899
Mech RF S	1.6388889	2.068139	0.79	0.4320
Mech RF R	2.9111111	0.827256	3.52	0.0010
Mech RF A	-0.611111	2.068139	-0.30	0.7689
Arm M	6.1388889	2.068139	2.97	0.0047
Arm CS	-0.722222	2.068139	-0.35	0.7285
Arm RF S	10	2.068139	4.84	<.0001
Arm RF R	8.9074074	0.68938	12.92	<.0001
Arm RF A	7.0833333	2.068139	3.42	0.0013
Avn M	1.5277778	2.068139	0.74	0.4637
Avn CS	-10.86111	2.068139	-5.25	<.0001
Avn RF S	3.25	2.068139	1.57	0.1226
Avn RF R	4.9166667	2.068139	2.38	0.0215
Avn RF A	-2.805556	2.068139	-1.36	0.1813
CAS RF S	0.3055556	2.068139	0.15	0.8832
CAS RF A	-3.472222	2.068139	-1.68	0.0997
B Eff	6.1111111	2.068139	2.95	0.0048

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	23	87781.78	3816.60	12.3932
Error	48	14782.00	307.96	Prob>F
C Total	71	102563.78		<.0001
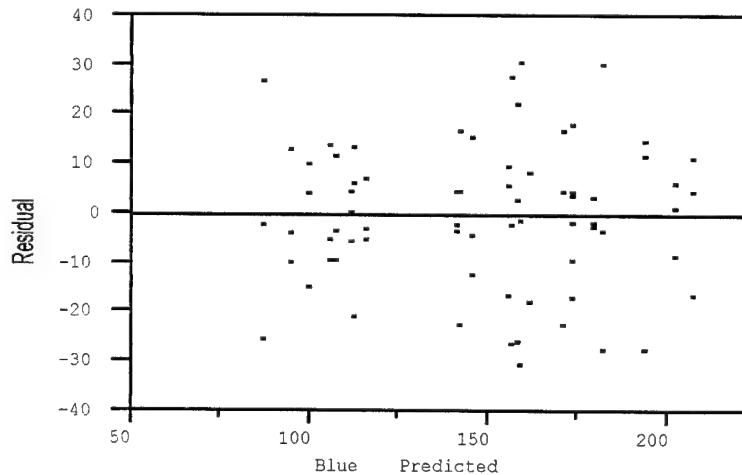
Percent Blue Screen

Summary of Fit

RSquare	0.839402
RSquare Adj	0.762449
Root Mean Square Error	6.695056
Mean of Response	58.50018
Observations (or Sum Wgts)	72

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	58.500179	0.78902	74.14	<.0001
Art M	-0.700458	0.78902	-0.89	0.3791
Art CS	0.4645959	0.78902	0.59	0.5587
Art RF S	-0.079222	0.78902	-0.10	0.9204
Art RF R	-1.178812	0.78902	-1.49	0.1417
Art RF A	0.9184657	0.78902	1.16	0.2502
Mech M	0.904299	0.78902	1.15	0.2574
Mech CS	-0.648735	0.78902	-0.82	0.4150
Mech RF S	1.3028991	0.78902	1.65	0.1052
Mech RF R	0.9807093	0.315608	3.11	0.0032
Mech RF A	1.2706316	0.78902	1.61	0.1139
Arm M	1.0249539	0.78902	1.30	0.2001
Arm CS	0.0624497	0.78902	0.08	0.9372
Arm RF S	3.4002367	0.78902	4.31	<.0001
Arm RF R	3.3150434	0.263007	12.60	<.0001
Arm RF A	1.5053198	0.78902	1.91	0.0624
Avn M	-0.419264	0.78902	-0.53	0.5976
Avn CS	-3.145199	0.78902	-3.99	0.0002
Avn RF S	2.1674957	0.78902	2.75	0.0084
Avn RF R	3.270775	0.78902	4.15	0.0001
Avn RF A	-0.835665	0.78902	-1.06	0.2948
CAS RF S	-0.02644	0.78902	-0.03	0.9734
CAS RF A	-0.663363	0.78902	-0.84	0.4047
B Eff	1.5130587	0.78902	1.92	0.0611

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	23	11245.532	488.936	10.9080
Error	48	2151.541	44.824	Prob>F
C Total	71	13397.073		<.0001

%Blue   Predicted

115
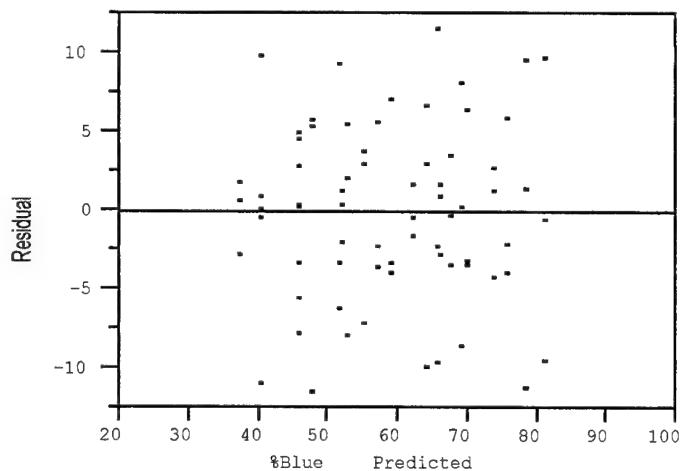
Appendix C: Remaining Factors

Remaining Significant Factors

Final Breakdown

Armor Factor		Low	Median	High
a	RF Strength	22	25	28
b	RF Range	2	3	4
	Aviation			
c	Combat Strength	58	65	72
d	RF Strength	31	35	39
e	RF Range	13	15	17

Resolution 5 Experiment

Experiment Design

TRIAL	a - Arm RF S	b - Arm RF R	c - Av CS	d - Av RF S	e - Av RF R
1	-1	-3	-1	-1	1
2	-1	-3	-1	1	-1
3	-1	-3	1	-1	-1
4	-1	-3	1	1	1
5	-1	3	-1	-1	-1
6	-1	3	-1	1	1
7	-1	3	1	-1	1
8	-1	3	1	1	-1
9	1	-3	-1	-1	-1
10	1	-3	-1	1	1
11	1	-3	1	-1	1
12	1	-3	1	1	-1
13	1	3	-1	-1	1
14	1	3	-1	1	-1
15	1	3	1	-1	-1
16	1	3	1	1	1
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0

Analysis of Factors--Run Summaries

Trial	Run	Blue	Red	Percent Blue	Trial	Run	Blue	Red	Percent Blue
1	1	127	128	49.8	11	51	124	135	47.9
	2	172	63	73.2		52	115	113	50.4
	3	113	136	45.4		53	119	134	47.0
	4	115	102	53.0		54	147	103	58.8
	5	132	104	55.9		55	137	131	51.1
2	6	115	161	41.7	12	56	124	105	54.1
	7	124	180	40.8		57	128	146	46.7
	8	138	84	62.2		58	103	128	44.6
	9	130	99	56.8		59	102	160	38.9
	10	132	128	50.8		60	150	117	56.2
3	11	129	130	49.8	13	61	211	50	80.8
	12	135	122	52.5		62	202	38	84.2
	13	120	143	45.6		63	224	27	89.2
	14	153	121	55.8		64	196	54	78.4
	15	132	116	53.2		65	187	52	78.2
4	16	110	128	46.2	14	66	169	67	71.6
	17	100	89	52.9		67	183	79	69.8
	18	129	74	63.5		68	164	62	72.6
	19	138	72	65.7		69	210	34	86.1
	20	151	112	57.4		70	193	77	71.5
5	21	154	84	64.7	15	71	176	104	62.9
	22	162	78	67.5		72	197	93	67.9
	23	211	58	78.4		73	167	73	69.6
	24	182	80	69.5		74	173	70	71.2
	25	186	42	81.6		75	185	69	72.8
6	26	179	41	81.4	16	76	203	56	78.4
	27	216	72	75.0		77	202	46	81.5
	28	201	24	89.3		78	201	12	94.4
	29	195	73	72.8		79	199	13	93.9
	30	206	40	83.7		80	205	27	88.4
7	31	203	34	85.7	17	81	172	74	69.9
	32	191	58	76.7	Center	82	140	58	70.7
	33	169	96	63.8		83	166	68	70.9
	34	185	66	73.7		84	182	53	77.4
	35	170	73	70.0		85	141	74	65.6
8	36	206	52	79.8	18	86	162	110	59.6
	37	132	132	50.0	Center	87	151	89	62.9
	38	191	65	74.6		88	175	105	62.5
	39	187	95	66.3		89	178	89	66.7
	40	190	86	68.8		90	166	107	60.8
9	41	113	132	46.1	19	91	150	88	63.0
	42	147	82	64.2	Center	92	175	90	66.0
	43	107	104	50.7		93	167	84	66.5
	44	140	131	51.7		94	159	69	69.7
	45	162	88	64.8		95	152	77	66.4
10	46	118	91	56.5	20	96	130	104	55.6
	47	170	89	65.6	Center	97	175	88	66.5
	48	132	119	52.6		98	178	104	63.1
	49	152	107	58.7		99	157	102	60.6
	50	112	79	58.6		100	185	69	72.8
					21	101	167	94	64.0
					Center	102	180	75	70.6
						103	152	78	66.1
						104	170	63	73.0
						105	134	104	56.3

Appendix D: Analysis of Factors

**Analysis of Factors
Response – Blue**

Normal Probability Plot
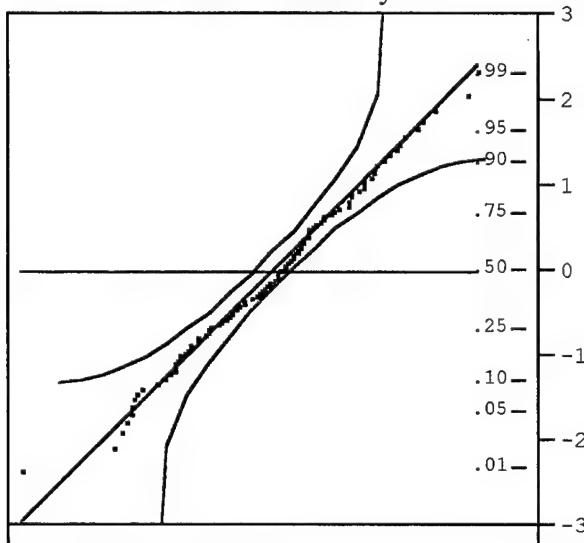

```
Summary of Fit

RSquare                            0.712016
RSquare Adj                        0.703462
Root Mean Square Error             17.05896
Mean of Response                   160.2286
Observations (or Sum Wgts)              105
```

```
Lack of Fit

Source         DF    Sum of Squares    Mean Square      F Ratio
Lack of Fit     1           178.354        178.354       0.6105
Pure Error    100         29213.460        292.135       Prob>F
Total Error   101         29391.814                      0.4364
                                                         Max RSq
                                                         0.7138
```

```
Parameter Estimates

Term                    Estimate    Std Error    t Ratio    Prob>|t|
Intercept              160.22857     1.664784      96.25     <.0001
b-Arm RF R             9.8583333      0.63575      15.51     <.0001
e-Av RF R                   4.45      1.90725       2.33     0.0216
b-Arm RF*e-Av RF       1.2416667      0.63575       1.95     0.0536
```

```
Analysis of Variance

Source       DF    Sum of Squares    Mean Square      F Ratio
Model         3          72668.70        24222.9      83.2379
Error       101          29391.81          291.0       Prob>F
C Total     104         102060.51                      <.0001
```

Analysis of Factors
Response – Percent Blue

Normal Probability Plot
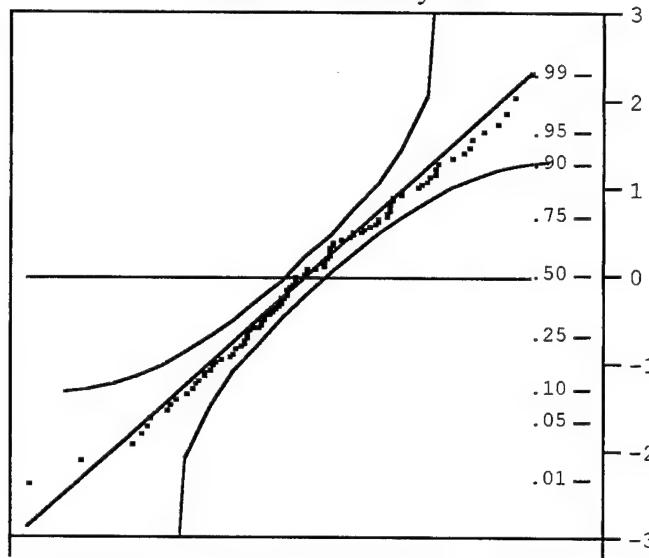

```
Summary of Fit

RSquare                          0.70119
RSquare Adj                      0.689237
Root Mean Square Error           6.989683
Mean of Response                 64.96717
Observations (or Sum Wgts)            105
```

```
Lack of Fit

Source          DF    Sum of Squares    Mean Square    F Ratio
Lack of Fit      4           73.6622        18.4156     0.3674
Pure Error      96         4811.9049        50.1240     Prob>F
Total Error    100         4885.5671                    0.8313
                                                        Max RSq
                                                        0.7057
```

```
Parameter Estimates

Term                   Estimate    Std Error    t Ratio    Prob>|t|
Intercept             64.967173     0.682123      95.24     <.0001
b-Arm RF R             3.7456642     0.26049      14.38      <.0001
c-Av CS               -1.45518       0.78147      -1.86      0.0655
e-Av RF R              3.565094      0.78147       4.56      <.0001
b-Arm RF*e-Av RF       0.4953962     0.26049       1.90      0.0601
```

```
Analysis of Variance

Source     DF    Sum of Squares    Mean Square    F Ratio
Model       4        11464.496        2866.12      58.6651
Error     100         4885.567          48.86      Prob>F
C Total   104        16350.064                      <.0001
```

119

Appendix E: Response Surfaces

For the Response -- Blue
(All coded +/-10%)
Best fit equation: $f(b, e) := 160 + 9.9 b + 4.5 e + 1.24 b \cdot e$

Where:

b = Arm RF Range (10%) e = Aviation RF Range

Blue ;   X axis = Armor RF Range   Y axis = Aviation RF Range
$b := 0$ $e := 0$ $f(b, e) = 160$
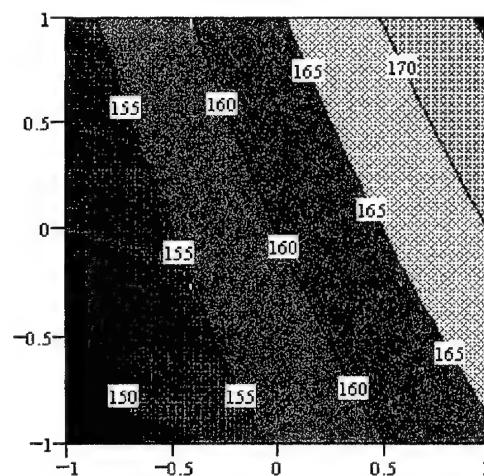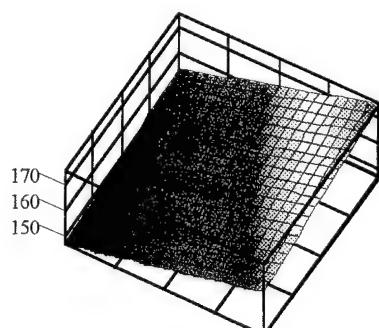
For the Response -- Percent Blue
(All coded +/- 10%)

Best fit equation: $f(b,c,e) := 65 + 3.75b - 1.46c + 3.57e + 0.50b \cdot e$

Where:

b = Arm RF Range   c = Aviation Combat Strength   e = Aviation RF Range

Percent Blue;   X axis = Armor RF Range   Y axis = Aviation Combat Strength

$b := 0 \quad c := 0 \quad e := 1 \quad f(b,c,e) = 68.57$
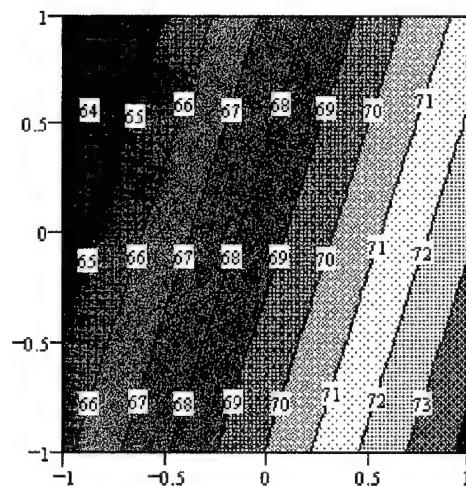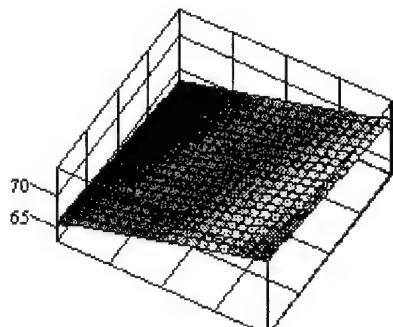
Percent Blue ;   X axis = Armor RF Range   Y axis = Aviation RF Range

$b := 0 \quad c := 1 \quad e := 0 \quad f(b, c, e) = 63.54$
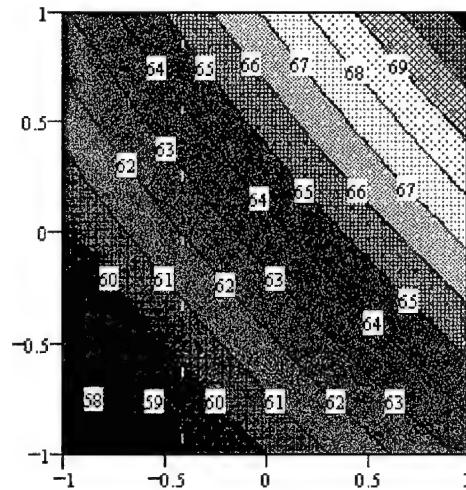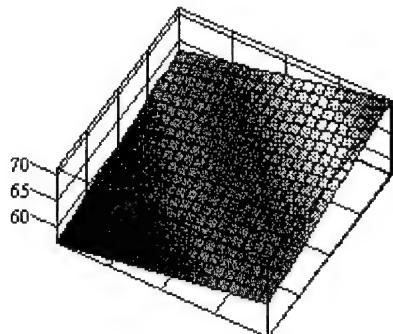
Percent Blue ;   X axis = Aviation Combat Strength   Y axis = Aviation RF Range

$b := 1 \quad c := 0 \quad e := 0 \quad f(b, c, e) = 68.75$
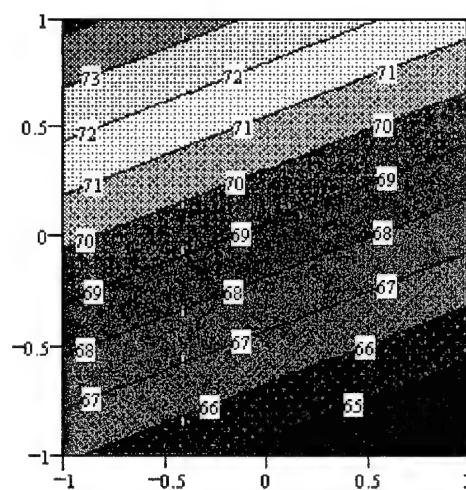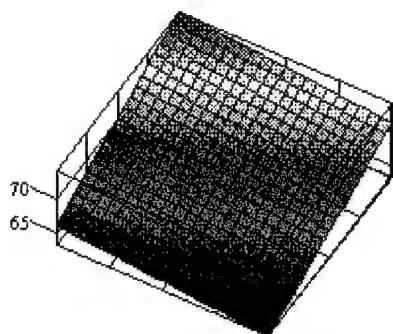
Appendix F: Validation

Full-Factorial Validation Experiment and Results

Trial	b - Arm RF R	c - Avn CS	e - Avn RF R
1	-3	-1	-1
2	-3	-1	1
3	-3	1	-1
4	-3	1	1
5	3	-1	-1
6	3	-1	1
7	3	1	-1
8	3	1	1
9	0	0	0
10	0	0	0
11	0	0	0
12	3	0	0
13	-3	0	0
14	0	1	0
15	0	-1	0
16	0	0	1
17	0	0	-1

Trial	Run	Blue	Red	Percent Blue	Trial	Run	Blue	Red	Percent Blue
1	1	142	127	52.8	9	41	172	74	69.9
	2	115	96	54.5		42	140	58	70.7
	3	125	105	54.3		43	166	68	70.9
	4	156	118	56.9		44	182	53	77.4
	5	102	97	51.3		45	141	74	65.6
2	6	178	108	62.2	10	46	162	110	59.6
	7	100	143	41.2		47	151	89	62.9
	8	117	127	48.0		48	175	105	62.5
	9	131	80	62.1		49	178	89	66.7
	10	134	94	58.8		50	166	107	60.8
3	11	161	108	59.9	11	51	150	88	63.0
	12	142	144	49.7		52	175	90	66.0
	13	131	103	56.0		53	167	84	66.5
	14	148	114	56.5		54	159	69	69.7
	15	115	120	48.9		55	152	77	66.4
4	16	95	110	46.3	12	56	212	37	85.1
	17	131	90	59.3		57	199	81	71.1
	18	95	132	41.9		58	180	68	72.6
	19	143	95	60.1		59	172	68	71.7
	20	128	61	67.7		60	176	71	71.3
5	21	204	65	75.8	13	61	126	99	56
	22	166	68	70.9		62	135	90	60
	23	187	94	66.5		63	139	109	56
	24	178	50	78.1		64	129	85	60.3
	25	196	61	76.3		65	122	140	46.6
6	26	190	56	77.2	14	66	154	86	64.2
	27	219	48	82.0		67	138	96	59
	28	181	51	78.0		68	160	70	69.6
	29	199	23	89.6		69	177	86	67.3
	30	168	52	76.4		70	136	114	54.4
7	31	177	76	70.0	15	71	172	49	77.8
	32	201	69	74.4		72	179	53	77.2
	33	184	74	71.3		73	147	99	59.8
	34	198	89	69.0		74	145	93	60.9
	35	193	65	74.8		75	203	61	76.9
8	36	190	81	70.1	16	76	141	87	61.8
	37	176	25	87.6		77	184	60	75.4
	38	212	43	83.1		78	110	105	51.2
	39	210	67	75.8		79	178	68	72.4
	40	168	57	74.7		80	181	77	70.2
					17	81	187	91	67.3
						82	179	86	67.5
						83	118	80	59.6
						84	143	85	62.7
						85	158	101	61

Validation
Full-Factorial Data

Response – Blue

Normal Probability Plot
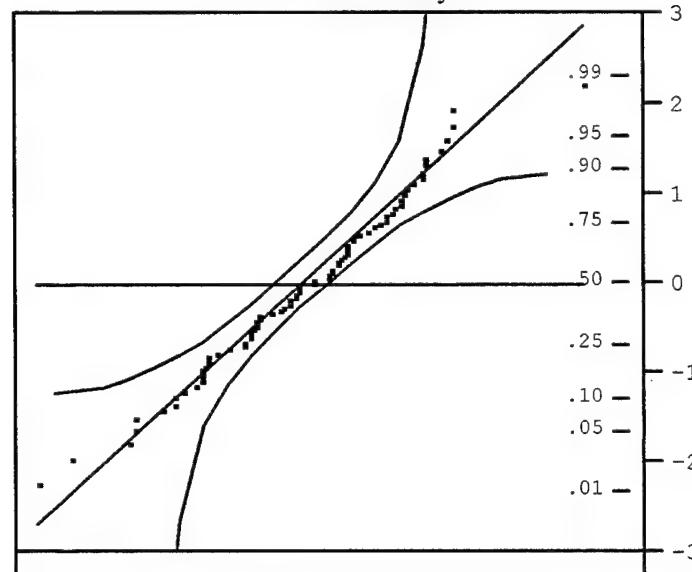
Summary of Fit	
RSquare	0.644871
RSquare Adj	0.630074
Root Mean Square Error	18.5789
Mean of Response	159.8026
Observations (or Sum Wgts)	76

Lack of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack of Fit	5	170.289	34.058	0.0924
Pure Error	67	24682.350	368.393	Prob>F
Total Error	72	24852.639		0.9932
				Max RSq
				0.6473

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	159.80263	2.131146	74.98	<.0001
b-Arm RF R	9.9733333	0.875818	11.39	<.0001
e-Avn RF R	-0.94	2.627453	-0.36	0.7216
b-Arm RF*e-Avn RF	0.95	0.979194	0.97	0.3352

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	45129.400	15043.1	43.5811
Error	72	24852.639	345.2	Prob>F
C Total	75	69982.039		<.0001

124

Validation
Full-Factorial Data

Response – Percent Blue

Normal Probability Plot
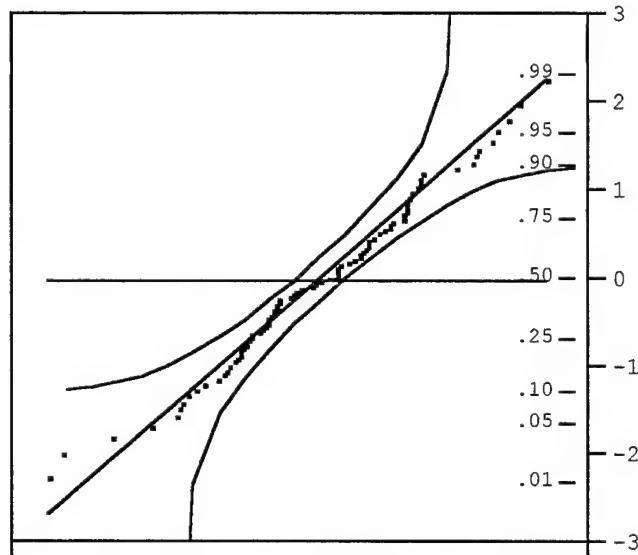

```
Summary of Fit
RSquare                                        0.658764
RSquare Adj                                    0.641702
Root Mean Square Error                         6.150654
Mean of Response                               65.59226
Observations (or Sum Wgts)                           85
```

```
Lack of Fit
Source          DF    Sum of Squares    Mean Square      F Ratio
Lack of Fit     10          170.2158        17.0216       0.4172
Pure Error      70         2856.2281        40.8033       Prob>F
Total Error     80         3026.4440                      0.9339
                                                          Max RSq
                                                          0.6780
```

```
Parameter Estimates
Term                     Estimate    Std Error    t Ratio    Prob>|t|
Intercept               65.592261     0.667132      98.32      <.0001
b-Arm RF R              3.5090242     0.289945      12.10      <.0001
c-Avn CS               -1.082454      0.869834      -1.24      0.2170
e-Avn RF R              1.7384296     0.869834       2.00      0.0491
b-Arm RF*e-Avn RF       0.5054451     0.324168       1.56      0.1229
```

```
Analysis of Variance
Source      DF    Sum of Squares    Mean Square    F Ratio
Model        4         5842.6258        1460.66    38.6105
Error       80         3026.4440          37.83    Prob>F
C Total     84         8869.0698                    <.0001
```

Plackett-Burman Validation Data

Run	b - Arm RF R	c - Avn CS	e - Avn RF R	Blue	Red	Percent Blue
1	3	1	1	219	22	90.87
2	3	1	1	212	51	80.61
3	3	1	1	191	76	71.54
4	3	1	1	129	83	60.85
5	3	1	1	158	69	69.60
6	3	1	1	190	55	77.55
7	3	1	1	179	53	77.16
8	3	1	1	178	103	63.35
9	3	1	1	165	130	55.93
10	3	1	1	140	79	63.93
11	3	1	1	166	108	60.58
12	3	1	1	162	100	61.83
13	3	1	-1	206	68	75.18
14	3	1	-1	167	73	69.58
15	3	1	-1	209	64	76.56
16	3	1	-1	159	114	58.24
17	3	1	-1	120	130	48.00
18	3	1	-1	147	102	59.04
19	3	-1	1	157	61	72.02
20	3	-1	1	192	68	73.85
21	3	-1	1	172	38	81.90
22	3	-1	1	204	28	87.93
23	3	-1	1	209	53	79.77
24	3	-1	1	194	95	67.13
25	3	-1	-1	213	66	76.34
26	3	-1	-1	179	89	66.79
27	3	-1	-1	155	78	66.52
28	3	-1	-1	183	89	67.28
29	3	-1	-1	177	99	64.13
30	3	-1	-1	178	72	71.20
31	3	-1	-1	162	80	66.94
32	3	-1	-1	181	75	70.70
33	3	-1	-1	133	113	54.07
34	3	-1	-1	170	135	55.74
35	3	-1	-1	144	117	55.17
36	3	-1	-1	170	87	66.15
37	-3	1	1	85	114	42.71
38	-3	1	1	110	115	48.89
39	-3	1	1	104	121	46.22
40	-3	1	1	119	103	53.60
41	-3	1	1	92	161	36.36
42	-3	1	1	126	111	53.16

Run	b - Arm RF R	c - Avn CS	e - Avn RF R	Blue	Red	Percent Blue
43	-3	1	-1	85	121	41.26
44	-3	1	-1	62	149	29.38
45	-3	1	-1	114	113	50.22
46	-3	1	-1	97	113	46.19
47	-3	1	-1	101	150	40.24
48	-3	1	-1	120	116	50.85
49	-3	1	-1	91	172	34.60
50	-3	1	-1	85	138	38.12
51	-3	1	-1	108	167	39.27
52	-3	1	-1	119	175	40.48
53	-3	1	-1	98	143	40.66
54	-3	1	-1	104	155	40.15
55	-3	-1	1	162	103	61.13
56	-3	-1	1	134	160	45.58
57	-3	-1	1	142	151	48.46
58	-3	-1	1	149	87	63.14
59	-3	-1	1	176	87	66.92
60	-3	-1	1	188	90	67.63
61	-3	-1	1	123	122	50.20
62	-3	-1	1	113	98	53.55
63	-3	-1	1	111	100	52.61
64	-3	-1	1	112	110	50.45
65	-3	-1	1	116	122	48.74
66	-3	-1	1	106	173	37.99
67	-3	-1	-1	131	114	53.47
68	-3	-1	-1	185	110	62.71
69	-3	-1	-1	155	128	54.77
70	-3	-1	-1	146	120	54.89
71	-3	-1	-1	138	169	44.95
72	-3	-1	-1	139	99	58.40
73	0	0	0	172	74	69.92
74	0	0	0	140	58	70.71
75	0	0	0	166	68	70.94
76	0	0	0	182	53	77.45
77	0	0	0	141	74	65.58
78	0	0	0	162	110	59.56
79	0	0	0	151	89	62.92
80	0	0	0	175	105	62.50
81	0	0	0	178	89	66.67
82	0	0	0	166	107	60.81
83	0	0	0	150	88	63.03
84	0	0	0	175	90	66.04
85	0	0	0	167	84	66.53
86	0	0	0	159	69	69.74
87	0	0	0	152	77	66.38
88	0	0	0	130	104	55.56
89	0	0	0	175	88	66.54
90	0	0	0	178	104	63.12
91	0	0	0	157	102	60.62
92	0	0	0	185	69	72.83

Validation
Plackett-Burman Data

Response – Blue

Normal Probability Plot
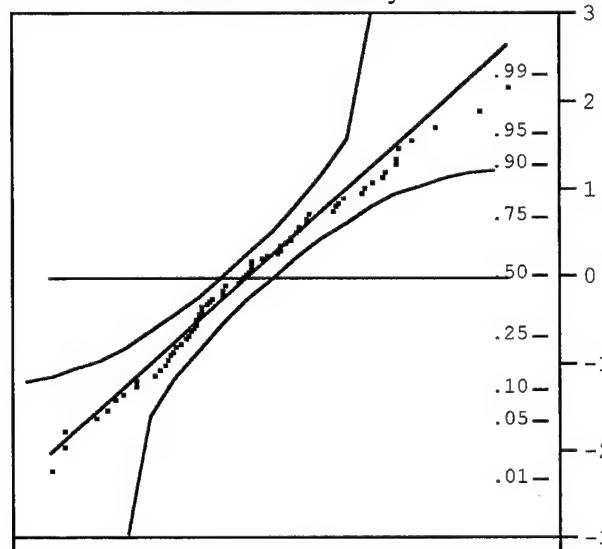

```
Summary of Fit
RSquare                           0.479786
RSquare Adj                       0.462052
Root Mean Square Error            25.59366
Mean of Response                   150.837
Observations (or Sum Wgts)              92
```

```
Lack of Fit
Source         DF    Sum of Squares    Mean Square    F Ratio
Lack of Fit     1         3811.827        3811.83      6.1605
Pure Error     87        53831.283         618.75      Prob>F
Total Error    88        57643.110                     0.0150
                                                       Max RSq
                                                        0.5142
```

```
Parameter Estimates
Term                   Estimate    Std Error    t Ratio    Prob>|t|
Intercept             150.83696     2.668323      56.53     <.0001
b-Arm RF R            8.9074074     1.005414       8.86     <.0001
e-Avn RF R            4.9166667     3.016242       1.63     0.1067
b-Arm RF*e-Avn RF      -0.12037     1.005414      -0.12     0.9050
```

```
Analysis of Variance
Source      DF    Sum of Squares    Mean Square    F Ratio
Model        3          53163.44        17721.1    27.0537
Error       88          57643.11          655.0    Prob>F
C Total     91         110806.55                   <.0001
```

Validation
Plackett-Burman Data

Response – Percent Blue

Normal Probability Plot
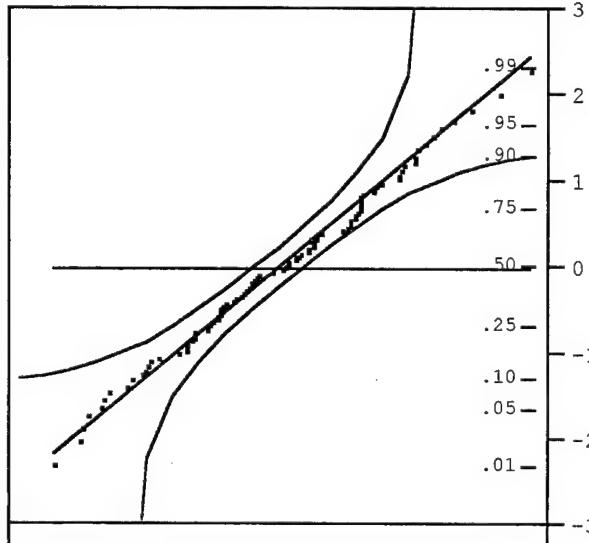

```
Summary of Fit

RSquare                                0.593473
RSquare Adj                            0.574783
Root Mean Square Error                 8.301257
Mean of Response                       60.10253
Observations (or Sum Wgts)                   92
```

```
Lack of Fit

Source          DF    Sum of Squares    Mean Square    F Ratio
Lack of Fit      4        1202.7047         300.676     5.2073
Pure Error      83        4792.5405          57.741    Prob>F
Total Error     87        5995.2452                     0.0009
                                                       Max RSq
                                                        0.6750
```

```
Parameter Estimates

Term                Estimate    Std Error    t Ratio    Prob>|t|
Intercept          60.102529    0.865466      69.45     <.0001
b-Arm RF R          3.3150434   0.326104      10.17     <.0001
c-Avn CS           -3.65302     1.037657      -3.52     0.0007
e-Avn RF R          3.270775    0.978312       3.34     0.0012
b-Arm RF*e-Avn RF   0.5078215   0.345886       1.47     0.1457
```

```
Analysis of Variance

Source      DF    Sum of Squares    Mean Square    F Ratio
Model        4        8752.242         2188.06     31.7520
Error       87        5995.245           68.91    Prob>F
C Total     91       14747.487                     <.0001
```

Predictive Analysis -- Full Factorial Plus Axial Data

b - Arm RFR	c - Avn CS	e - Avn RFR	Blue	% Blue	Predicted Blue	Predicted % Blue
-3	-1	-1	142	52.8	130	53.1
-3	-1	-1	115	54.5	130	53.1
-3	-1	-1	125	54.3	130	53.1
-3	-1	-1	156	56.9	130	53.1
-3	-1	-1	102	51.3	130	53.1
-3	-1	1	178	62.2	131	57.3
-3	-1	1	100	41.2	131	57.3
-3	-1	1	117	48.0	131	57.3
-3	-1	1	131	62.1	131	57.3
-3	-1	1	134	58.8	131	57.3
-3	1	-1	161	59.9	130	50.2
-3	1	-1	142	49.7	130	50.2
-3	1	-1	131	56.0	130	50.2
-3	1	-1	148	56.5	130	50.2
-3	1	-1	115	48.9	130	50.2
-3	1	1	95	46.3	131	54.4
-3	1	1	131	59.3	131	54.4
-3	1	1	95	41.9	131	54.4
-3	1	1	143	60.1	131	54.4
-3	1	1	128	67.7	131	54.4
3	-1	-1	204	75.8	181	72.6
3	-1	-1	166	70.9	181	72.6
3	-1	-1	187	66.5	181	72.6
3	-1	-1	178	78.1	181	72.6
3	-1	-1	196	76.3	181	72.6
3	-1	1	190	77.2	198	82.8
3	-1	1	219	82.0	198	82.8
3	-1	1	181	78.0	198	82.8
3	-1	1	199	89.6	198	82.8
3	-1	1	168	76.4	198	82.8
3	1	-1	177	70.0	181	69.7
3	1	-1	201	74.4	181	69.7
3	1	-1	184	71.3	181	69.7
3	1	-1	198	69.0	181	69.7
3	1	-1	193	74.8	181	69.7
3	1	1	190	70.1	198	79.9
3	1	1	176	87.6	198	79.9
3	1	1	212	83.1	198	79.9
3	1	1	210	75.8	198	79.9
3	1	1	168	74.7	198	79.9
0	0	0	172	69.9	160	65
0	0	0	140	70.7	160	65
0	0	0	166	70.9	160	65
0	0	0	182	77.4	160	65
0	0	0	141	65.6	160	65
0	0	0	162	59.6	160	65
0	0	0	151	62.9	160	65
0	0	0	175	62.5	160	65
0	0	0	178	66.7	160	65
0	0	0	166	60.8	160	65
0	0	0	150	63.0	160	65
0	0	0	175	66.0	160	65
0	0	0	167	66.5	160	65
0	0	0	159	69.7	160	65
0	0	0	152	66.4	160	65
3	0	0	212	85.1	190	76.3
3	0	0	199	71.1	190	76.3
3	0	0	180	72.6	190	76.3
3	0	0	172	71.7	190	76.3
3	0	0	176	71.3	190	76.3
-3	0	0	126	56	130	53.75
-3	0	0	135	60	130	53.75
-3	0	0	139	56	130	53.75
-3	0	0	129	60.3	130	53.75
-3	0	0	122	46.6	130	53.75
0	1	0	154	64.2	160	63.5
0	1	0	138	59	160	63.5
0	1	0	160	69.6	160	63.5
0	1	0	177	67.3	160	63.5
0	1	0	136	54.4	160	63.5
0	-1	0	172	77.8	160	66.5
0	-1	0	179	77.2	160	66.5
0	-1	0	147	59.8	160	66.5
0	-1	0	145	60.9	160	66.5
0	-1	0	203	76.9	160	66.5
0	0	1	141	61.8	164.5	68.6
0	0	1	184	75.4	164.5	68.6
0	0	1	110	51.2	164.5	68.6
0	0	1	178	72.4	164.5	68.6
0	0	1	181	70.2	164.5	68.6
0	0	-1	187	67.3	155.5	61.4
0	0	-1	179	67.5	155.5	61.4
0	0	-1	118	59.6	155.5	61.4
0	0	-1	143	62.7	155.5	61.4
0	0	-1	158	61	155.5	61.4

	Blue	**% Blue**
Relative Error	1.61%	-0.12%
RMSE	18.94	6.19

Predictive Analysis -- Plackett-Burman Screening Data

b - Arm R F R	c - Avn C S	e - Avn R F R	Blue	% Blue	Predicted Blue	Predicted % Blue
3	1	1	219	90.87137	198	79.9
3	1	1	212	80.60837	198	79.9
3	1	1	191	71.53558	198	79.9
3	1	1	129	60.84906	198	79.9
3	1	1	158	69.60352	198	79.9
3	1	1	190	77.55102	198	79.9
3	1	1	179	77.15517	198	79.9
3	1	1	178	63.3452	198	79.9
3	1	1	165	55.9322	198	79.9
3	1	1	140	63.92694	198	79.9
3	1	1	166	60.58394	198	79.9
3	1	1	162	61.83206	198	79.9
3	1	-1	206	75.18248	181	69.7
3	1	-1	167	69.58333	181	69.7
3	1	-1	209	76.55678	181	69.7
3	1	-1	159	58.24176	181	69.7
3	1	-1	120	48	181	69.7
3	1	-1	147	59.03614	181	69.7
3	-1	1	157	72.01835	198	82.8
3	-1	1	192	73.84615	198	82.8
3	-1	1	172	81.90476	198	82.8
3	-1	1	204	87.93103	198	82.8
3	-1	1	209	79.77099	198	82.8
3	-1	1	194	67.12803	198	82.8
3	-1	-1	213	76.34409	181	72.6
3	-1	-1	179	66.79104	181	72.6
3	-1	-1	155	66.52361	181	72.6
3	-1	-1	183	67.27941	181	72.6
3	-1	-1	177	64.13043	181	72.6
3	-1	-1	178	71.2	181	72.6
3	-1	-1	162	66.94215	181	72.6
3	-1	-1	181	70.70313	181	72.6
3	-1	-1	133	54.06504	181	72.6
3	-1	-1	170	55.7377	181	72.6
3	-1	-1	144	55.17241	181	72.6
3	-1	-1	170	66.14786	181	72.6
-3	1	1	85	42.71357	131	54.4
-3	1	1	110	48.88889	131	54.4
-3	1	1	104	46.22222	131	54.4
-3	1	1	119	53.6036	131	54.4
-3	1	1	92	36.36364	131	54.4
-3	1	1	126	53.16456	131	54.4
-3	1	-1	85	41.26214	130	50.2
-3	1	-1	62	29.38389	130	50.2
-3	1	-1	114	50.22026	130	50.2
-3	1	-1	97	46.19048	130	50.2
-3	1	-1	101	40.23904	130	50.2
-3	1	-1	120	50.84746	130	50.2
-3	1	-1	91	34.60076	130	50.2
-3	1	-1	85	38.11659	130	50.2
-3	1	-1	108	39.27273	130	50.2
-3	1	-1	119	40.47619	130	50.2
-3	1	-1	98	40.6639	130	50.2
-3	1	-1	104	40.15444	130	50.2
-3	-1	1	162	61.13208	131	57.3
-3	-1	1	134	45.57823	131	57.3
-3	-1	1	142	48.46416	131	57.3
-3	-1	1	149	63.13559	131	57.3
-3	-1	1	176	66.92015	131	57.3
-3	-1	1	188	67.6259	131	57.3
-3	-1	1	123	50.20408	131	57.3
-3	-1	1	113	53.5545	131	57.3
-3	-1	1	111	52.60664	131	57.3
-3	-1	1	112	50.45045	131	57.3
-3	-1	1	116	48.7395	131	57.3
-3	-1	1	106	37.99283	131	57.3
-3	-1	-1	131	53.46939	130	53.1
-3	-1	-1	185	62.71186	130	53.1
-3	-1	-1	155	54.77032	130	53.1
-3	-1	-1	146	54.88722	130	53.1
-3	-1	-1	138	44.95114	130	53.1
-3	-1	-1	139	58.40336	130	53.1
0	0	0	172	69.9187	160	65
0	0	0	140	70.70707	160	65
0	0	0	166	70.94017	160	65
0	0	0	182	77.44681	160	65
0	0	0	141	65.5814	160	65
0	0	0	162	59.55882	160	65
0	0	0	151	62.91667	160	65
0	0	0	175	62.5	160	65
0	0	0	178	66.66667	160	65
0	0	0	166	60.80586	160	65
0	0	0	150	63.02521	160	65
0	0	0	175	66.03774	160	65
0	0	0	167	66.53388	160	65
0	0	0	159	69.73684	160	65
0	0	0	152	66.37555	160	65
0	0	0	130	55.55556	160	65
0	0	0	175	68.53992	160	65
0	0	0	178	63.12057	160	65
0	0	0	157	60.61776	160	65
0	0	0	185	72.83465	160	65

	Blue	% Blue
Relative Error	10%	11%
RMSE	19.51	6.26

Bibliography

Armed Forces Staff College. The Joint Staff Officer's Guide 1993. AFSC Pub 1. Norfolk: National Defense University (1993).

Battilega, John A. and Judith K. Grange (Editors), The Military Applications of Modeling, AFIT Press, Wright Patterson AFB OH (1984).

Box, George E., Norman R. Draper, Empirical Model-Building and Response Surfaces, New York: Wiley (1987).

Brewer, Garry D. and Martin Shubik, The War Game, Cambridge: Harvard University Press (1979).

Department of the Army. Operations. FM 100-5. Washington: HQ USA, 14 June 1993.

Draper, N.R. and H. Smith, Applied Regression Analysis 2nd Edition New York: John Wiley & Sons (1981).

Dunnigan, James F. The Complete Wargames Handbook – How to Play, Design & Find Them (Revised Edition), New York: William Morrow (1992).

Dupuy, T.N. Numbers, Predictions and War, Fairfax, Virginia: HERO Books (1985).

Farmer, Michael Ryan, The Effects of Changing Force Structure on THUNDER Output, MS Thesis, AFIT/GOA/ENS/96M-01. School of Engineering, Air Force Institute of Technology (AU), Wright Patterson AFB OH, March 1996.

Forsythe, Steven L. Optimization of the Air Apportionment in a TAC THUNDER Scenario Using Response Surface Methodology. MS Thesis, AFIT/GST/ENS/94M-02. School of Engineering, Air Force Institute of Technology (AU), Wright Patterson AFB OH, March 1994.

Ghelber, Craig S. and Charles A. Haley, A Methodology for Validation of Complex Multi-variable Military Computerized Models, MS Thesis, AFIT/GOR/MA/80D-3. School of Engineering, Air Force Institute of Technology (AU), Wright Patterson AFB OH, December 1980.

Goldfarb, Alan S. and others, "Sensitivity Analysis Using Discrete Simulation Models", The Proceedings of the 1988 Summer Computer Simulation Conference, Simulation Councils Ltd. 689-693.

Hartman, James K., Lecture Notes in Aggregated Combat Modeling. Department of Operations Research, Naval Postgraduate School, Monterey, CA (1985), as edited by Lt Col Jack Kloeber (USA) and Lt Col Jack Jackson (USAF), Air Force Institute of Technology, (1996)

Hillestad, Richard, Louis Moore, and Bart Bennett, "Modeling for Campaign Analysis: Lessons for the Next Generation of Models." Unpublished Report DRR-1088-AF RAND September 1995 (Draft).

Iman, Ronald L., Jon C. Helton, and James E. Campbell, "An Approach to Sensitivity Analysis of Computer Models" -- Parts I and II, *Journal of Quality Technology*, Vol 13, No 3&4, July 1981, 174-183 and October 1981, 232-240.

Kleijnen, Jack P.C., Statistical Techniques in Simulation (Part II), New York: Dekker (1975).

Law, Averill M. and W. David Kelton, Simulation Modeling & Analysis 2nd Edition New York: McGraw-Hill (1991).

Mendenhall, William, Introduction to Linear models and The Design and Analysis of Experiments, Belmont CA: Wadsworth (1968).

Mendenhall, William, Dennis D. Wackerly and Richard L. Scheaffer, Mathematical Statistics With Applications (4th Edition), Belmont, CA: Duxbury (1990)

Miser, Hugh J. and Edward S. Quade, Handbook of Systems Analysis – Overview of Uses, Procedures, Applications, and Practice, New York: North-Holland (1985).

Myers, David A. Response Surface Methodology as a Sensitivity Tool in Decision Analysis, MS Thesis, AFIT/GOA/ENS/95M-05. School of Engineering, Air Force Institute of Technology (AU), Wright Patterson AFB OH, March 1995.

Myers, Raymond H. and Douglas C. Montgomery, Response Surface Methodology New York: John Wiley & Sons (1995).

Myers, Raymond H., Response Surface Methodology, Published by Author (1976).

Naylor, Thomas H. (Editor), The Design of Computer Simulation Experiments, Durham, NC: Duke University Press (1969).

New World Computing™ Inc., *Empire II: The Art of War*™ User Manual (1995).

Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman, <u>Applied Linear Statistical Models</u> 4th Edition, Chicago: Irwin (1996).

Stone, Matthew A. <u>Decreasing Nonconformance of Parts in the Air Force Supply System</u>, MS Thesis, AFIT/GOR/MA/88D-6. School of Engineering, Air Force Institute of Technology (AU), Wright Patterson AFB OH, December 1988.

Taguchi, G. <u>Orthogonal Arrays and Linear Graphs</u>, Dearborn MI: ASI Press (1987).

Wu, C.F.J., Youhi Chen, "A Graph-Aided Method for Planning Two-Level Experiments When Certain Interactions are Important", <u>*Technometrics*</u>, Vol. 34, No. 2, May 1992, 162-175

Vita

Major William Berg was born on March 4, 1961 in Albuquerque, New Mexico. In 1979, he graduated from Sandia High School in Albuquerque, then enrolled at New Mexico State University in Las Cruces on a four-year Air Force ROTC scholarship. In May, 1983, Major Berg received a Bachelor of Science in Civil Engineering, and was commissioned as a Distinguished Graduate. He attended Undergraduate Navigator Training at Mather AFB, California where he finished as a Distinguished Graduate. After flying three years as a Weapon Systems Officer in the F-111D at Cannon AFB, New Mexico he was selected to attend Undergraduate Pilot Training with the Euro-NATO Joint Jet Pilot Training (ENJJPT) program at Sheppard AFB, Texas. Major Berg received the Air Training Command Commander's Cup for graduating first in his class. After upgrade training at Luke AFB, Arizona, he flew the F-16 (A and C), at Homestead AFB, Florida. In 1992, Major Berg was selected to fly the F-117A at Holloman AFB, New Mexico. While at Holloman, he served as a Flight Commander and F-117A Instructor Pilot. In the summer of 1995, he entered the Operational Analysis program at the Air Force Institute of Technology. After receiving his degree, Major Berg will be assigned to the Combat Analysis unit of the United States Central Command (USCENTCOM) at MacDill AFB, Florida.

Permanent Address: 2901 San Pablo NE
                      Albuquerque, New Mexico 87110

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1997	Master's Thesis

4. TITLE AND SUBTITLE

SENSITIVITY ANALYSIS OF A COMBAT SIMULATION USING RESPONSE SURFACE METHODOLOGY

5. FUNDING NUMBERS

6. AUTHOR(S)

William J. Berg, Major, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
2750 P Street
WPAFB, OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This thesis provides a methodology to analyze the sensitivity of a combat simulation relative to the capabilities of the modeled weapon systems. Response Surface Methodology (RSM) was the primary technique used for analysis. A ground battle scenario was developed with a commercial PC wargame *Empire II: The Art of War*™. In the computer verses computer mode, the wargame was transformed into a 'combat simulation'. A screening experiment was used to determine which of an original 23 weapon factors were actually significant to this scenario. After analyzing the remaining five factors, only three ended up as truly significant to the battle outcome. The three factors are used in a regression equation which predicts the outcome of the overall simulation scenario. The sensitivity of the simulation outcome is reflected by the coefficients of each factor. The regression model was then validated against independent simulation trials.

The method proved useful in analyzing the underlying behavior of a combat simulation. Combat analysts must isolate those weapon capabilities which have the greatest influence on the simulation model. Modeling funds can be concentrated on those factors. With accurate modeling and timely updates of those capabilities, the simulation model will earn more credibility as a decision aid.

14. SUBJECT TERMS

Response Surface Methodology, RSM, Combat Simulation, Wargame, Simulation, Regression Analysis, Sensitivity Analysis

15. NUMBER OF PAGES

145

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL